DOCUMENT RESUME

ED 118 660 95 UD 015 612

AUTHOR	Cooley, William W.; Leinhardt, Gaea
TITLE	Design for the Individualized Instruction Study: A Study of the Effectiveness of Individualized Instruction in the Teaching of Reading and Mathematics in Compensatory Education Programs. Final Report.
INSTITUTION	Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Oct 75
NOTE	271p.
EDRS PRICE	MF-$0.83 HC-$14.05 Plus Postage
DESCRIPTORS	Academic Achievement; *Compensatory Education Programs; *Educational Assessment; Educational Research; Enrichment; Evaluation Criteria; *Evaluation Methods; Federal Programs; Individual Instruction; *Individualized Instruction; Individualized Reading; Mathematics Instruction; Minority Group Children; Program Evaluation; Reading Achievement; Reading Instruction; Reading Programs; *Research Design; Research Methodology

ABSTRACT

This report presents a detailed rationale and a plan for a study of the effectiveness of materials and procedures for meeting the educational needs of individual children in compensatory education programs. An introductory section addresses an overview of the policy questions to which the results of the individualized instruction study will be relevant, an outline of the distinguishing features of the proposed approach, and a discussion of alternative designs that will serve to further clarify the design that is recommended. Other sections include foci on defining and assessing program and support variables, assessing outcomes, sampling and securing the cooperation of schools, collecting data, reducing and analyzing data, project organization, staffing, schedule and references. Appendices include in-class interviewing, analyzing curricula, behavioral analysis of curricula, videotaping, and test reviews. The outcome of primary interest in this study is stated to be achievement in reading and mathematics, other outcomes also having been considered for inclusion. The assessment section describes the outcome measure that is recommended for assessing student achievement and the rationale for its selection. In addition, measurement of program effects on noncognitive student outcomes and on the total classroom environment is discussed. (Author/AM)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

15612

ED118660

Design for the Individualized Instruction Study: A Study of the Effectiveness of Individualized Instruction in the Teaching of Reading and Mathematics in Compensatory Education Programs

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Final Report

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: UD | SP
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

William W. Cooley and Gaea Leinhardt
Co-Project Coordinators

Learning Research and Development Center
University of Pittsburgh

October 1975

UD 015 612

Acknowledgments

Although we had overall responsibility for this design study, we want to acknowledge the extensive help that we have received from many others.

First of all, the administrative and editorial assistance provided by Janet McGrail was an essential ingredient. Without her, the final report would probably never have been completed.

Important in the formulation of the general approach taken in the design was a weekly seminar attended by several LRDC Research Associates. Through these seminars, we obtained many important ideas from R. Tony Eichelberger, Robert Glaser, J. Ernest Harrison, James G. Holland, David F. Lancy, C. M. Lindvall, and Lauren B. Resnick, as well as constructive criticism of our own ideas.

Several Research Assistants helped with various tasks, such as instrument development and test reviews. These individuals include: David D. R. DeYoung, James L. DiCostanzo, Sundari Krishnamurthy, Andrea M. Seewald, and Allyson Walker.

Major typing and proofreading efforts were provided by Jane Komaromy and Terri Komar, who graciously and competently prepared the many versions of this report.

Numerous useful suggestions were received from several consultants who reacted to the September draft of the report. Participating as consultants were: John B. Carroll, Robert M. Gagné, Annegret Harnischfeger, Susan M. Markle, and David E. Wiley. Special mention should be made of Paul R. Lohnes, who provided draft text for sections of this report.

William W. Cooley
Gaea Leinhardt

Table of Contents

Page

Section

1 Introduction 1

2 Defining and Assessing Program and Support Variables. 11

3 Assessing Outcomes. 31

4 Sampling and Securing the Cooperation of Schools. 41

5 Collecting Data 51

6 Reducing and Analyzing Data 59

7 Project Organization, Staffing, and Schedule. 80

References. 89

Appendix

A In-Class Interviewing A-1

B-1 Analyzing Curricula B-1

B-2 Behavioral Analysis of Curricula B-53

C Videotaping. C-1

D Test Reviews D-1

Section 1

Introduction

Congress has directed the National Institute of Education (NIE) to perform a general examination of compensatory education programs, with the specific instruction to conduct a detailed study of the effectiveness of materials and procedures for meeting the educational needs of individual children. This report presents a detailed rationale and plan for the conduct of such a study. Included in this introductory section of the report are an overview of the policy questions to which the results of the Individualized Instruction Study will be relevant, an outline of the distinguishing features of our proposed approach, and a discussion of alternative designs that will serve to further clarify the design that we are recommending.

Policy Questions

The findings of the Individualized Instruction Study will be relevant to policy questions that have been of major concern to Congress, to NIE, and to others interested in education. The overall question concerns the success with which various educational approaches are compensating for children's initial educational disadvantage by meeting their individual needs. Related to this question are two major issues. The first concerns the ways in which compensatory education funds can be most effectively utilized. Should funds be provided for supplemental services and materials for disadvantaged children, should they be provided for specific programs in which only students labeled as educationally disadvantaged can participate, or should an alternative allocation procedure be adopted? The second issue is whether or not compensatory education funds should be provided at all. If it is true that schools don't affect student learning, as several large-scale studies have suggested, then why continue to spend more and more money on educational programs, compensatory or otherwise. The design that we are recommending has been formulated with these issues in mind.

To better understand the relevance of the study proposed here to the first of these policy issues, it is useful to review the procedures followed in Title I of the Elementary and Secondary Education Act, as revealed by our informal examination of programs in Pennsylvania and New York. Title I is the most massive Federal effort for the support of compensatory education. Within a school district, the allocation process begins with the collection of information on the socioeconomic status of families living in the attendance area for a particular school. That school's eligibility for Title I funds is a function of the number of children from low income families within its attendance area in relation to the number of economically disadvantaged children from other attendance areas in the school district. Generally, any school having a percentage of low income children equal to or exceeding the district-wide percentage is eligible for Title I funds.

However, within a school designated as eligible, only those children who are defined as educationally disadvantaged may actually benefit from Title I funds.* Educationally disadvantaged children are those who need special assistance in order to raise their level of educational attainment to the level appropriate for children of their age. This group includes children whose needs for assistance result from poverty, neglect, delinquency, or cultural or linguistic isolation from the community at large. Most frequently, educationally disadvantaged is operationally defined as a certain grade level behind one's classmates on standardized achievement tests administered by the school system in the fall.

---

*Another important policy issue is whether funding, even if supplemental, should be based on economic disadvantage or educational disadvantage. It would seem that this question can be settled most easily by establishing the relationship between economic indicators and educational achievement indicators. Reviewing data for two urban school systems (Pittsburgh and Minneapolis), we found that correlations between two economic indicators (percentage of families receiving Aid for Dependent Children funds and percentage of families earning below $2,000) correlated in the range .75 to .95 with the percentage of children in a school scoring at least one-half year below grade level, using the school as the unit of analysis. Thus, schools would receive essentially the same support whether funding is based on economic or achievement indicators. Economic indicators seem to be the preferred political solution because the distribution of funds can be estimated in advance and because more members of Congress are willing to view Title I support as a poverty measure. Also, distribution based on economic indicators eliminates the negative incentive to improve achievement, which is built into allocation procedures that are keyed to low achievement.

Title I funds are administered by local education agencies (LEAs). At present, the LEAs must demonstrate that the funds are indeed being used for programs that are supplemental to standard classroom instruction. Title I programs in the lower elementary school often involve the hiring of a special tutor who works with those children labeled as disadvantaged. Tutoring generally occurs in a corner of the classroom or in a separate room while the regular teacher is teaching something else to the rest of the class--instruction that the disadvantaged children would not have the opportunity to benefit from. The confusions and inequities that the present allocation process produces are even more exaggerated when one considers a classroom where a majority of the students in the room are eligible for Title I funds as defined by their low achievement scores. How does one supplement a classroom program with Title I funds when 70 percent of the students in the room are eligible? One approach is for the school to segregate students on the basis of achievement scores and have an entire classroom of all eligibles.

It seems that it would be far more effective if Title I funding could be used to subsidize a strong basic program for all children in schools that include some specific percentage of children designated as eligible for special assistance. A basic program that is adaptive to individual differences in all students would probably be more educationally sound than the programs currently supported by Title I that are restricted to eligible children. Secondly, an individualized program could eliminate the need to brand any particular group of students as disadvantaged. One negative factor in compensatory programs is that being labeled disadvantaged often produces a disadvantage (e.g., White et al., 1973). Third, it would be easier to monitor a core program since only budgets and allocations would need to be examined. For example, it would not be necessary, as it is under current guidelines, to make sure that a classroom aide provided through Title I support helps only those students in the classroom who are designated as eligible.

A related policy issue to which the results of the Individualized Instruction Study will also be relevant concerns the extent to which low performance in basic skills such as reading and mathematics can be affected by special educational programs. Major studies such as the Equality of Educational Opportunity survey (Coleman et al., 1966) and the later syntheses of this and related efforts (e.g., Averch et al., 1972; Jencks et al., 1972) are commonly cited as evidence that innovative educational programs do not significantly improve student performance or, at least, that different programs do not make a difference in what children learn. The results of the Individualized Instruction Study can possibly be useful in challenging this trend, since what happens at the classroom level is to be investigated. Prior research used school variables that were too far removed from the educational program actually being implemented in the classroom to have much effect on what students in that classroom learned. Differences in variables that might have made a difference at the classroom level often "washed out" at the school level, leading to the conclusion that school differences don't make a difference. Allowing this conclusion to stand unchallenged would be to give up on improving education.

Given the confusion of the current allocation procedure and the enormous cost of compensatory education, Congress has requested detailed information on the administration of compensatory programs, fund allocation systems, and student development. NIE has chosen to obtain the requested information by funding six major studies, four of which relate to student development: a nationally representative survey of ongoing programs, a synthesis of results to date, a review of alternative designs for compensatory education, and an in-depth review of the effects of individualized instruction. The results of this latter study, for which we are proposing a design, should not only clarify for Congress what individualization can and cannot do for education, but also provide a detailed picture of compensatory education at the classroom rather than the school level. The contractor should be aware of the context in which the Individualized Instruction Study is being undertaken and the magnitude of its potential.

Overview of Design Features

The design that we are recommending incorporates several distinguishing features. These features concern program and implementation measures, outcome measures, sampling procedures, and analysis procedures. In addition, a special study is proposed to contrast the effectiveness of instructional settings in which the education of compensatory children can occur. Each of these features is described briefly below.

Program and implementation variables. Our design assumes that classrooms differ in the manner in which and degree to which they are adaptive to individual differences in children. Classroom practices can vary in how they provide for differences in the knowledges and abilities with which children begin a school year, in how they respond to differences in interest and motivation, in how they adapt to different rates of learning, in whether or not they recognize and deal with cultural differences, and in whether or not they allow for a variety of educational goals. This variation is the result of: (1) differences in the instructional programs being used in classrooms, and (2) differences in the way in which teachers implement any given program. The design that we are recommending includes a set of variables that will be used in assessing the actual processes that are taking place in classrooms and in determining the relative effectiveness of these processes. The model of classroom processes that guided the formulation of our design is a model for explaining the amount of learning that has taken place in a classroom during the school year. The assumption is that variation in learning is a function of the processes that are operating in that classroom and the initial abilities of children.

We have not formulated a set of criteria to determine whether or not a particular school or classroom is operating an individualized or standardized program, and then a second set of criteria to determine how well the school or classroom has implemented that program. Instead, we have defined a set of dimensions along which all classrooms differ and in terms of which they will be assessed. The dimensions reflect both program design differences and implementation variation. Those classroom dimensions that are related to aspects of individualization are referred to as program variables. Other classroom dimensions that are related to learning but not to individualization are viewed as support variables.

Outcome measures. We propose that commercially available achievement tests in reading and mathematics be used in the Individualized Instruction Study. A test battery is recommended that is superior to any battery that could be built in the time available to the contractor, certainly in terms of reliability and the elimination of racially biasing items, and probably also in terms of content validity.

Sampling procedures. The primary objective of our design is to identify effective mechanisms for improving educational programs, not to evaluate Title I or any other compensatory programs. Too little is learned from contrasts that assume that programs are unique entities that can be implemented exactly as they were designed. Therefore, the sampling scheme emphasizes the need to achieve variance in classroom processes, rather than representativeness of present compensatory practices. It also assumes that the study can best be accomplished by an intensive examination of classroom practices in a restricted geographic region, and that little would be gained from the extra expense of national representation.

Analysis procedures. The analytic procedure that we propose is based on the recognition that classroom practices will occur in uncontrollable combinations in the field. What is needed in such situations is a technique for sorting out what can be uniquely attributed to particular practices, and what is due to combinations of practices or relationships between practices and initial abilities of children.

One specific technique that we recommend is commonality analysis, which is a variation of regression analysis that allows one to distinguish between unique effects and effects that are common to two or more of the predictor variables, where the predictors are both classroom process dimensions and children's initial abilities. This technique will provide the basis for inferences about the extent to which differences in classroom processes produce differences in student achievement, the extent to which classroom process effects are due to methods of individualization, and the relative effectiveness of different individualizing mechanisms.

Instructional setting contrast. The main thrust of the study will be to identify effective classroom practices, particularly those related to individual differences in children, with the objective of showing how the goals of compensatory education can be achieved without the need to label specific children as disadvantaged. To complete the study, it will be necessary to contrast the more frequently used Title I practice, which is to provide special tutors for eligible children with the alternative of providing for these children in classrooms that adapt to differences among all children. Section 6 of this report outlines an approach to this contrast of instructional settings.

Alternative Designs

The design that we are suggesting can, of course, be modified in many ways. For example, a broader geographic distribution of classrooms could be incorporated into the design so as to reduce the likelihood that regional differences could affect the outcomes of the study. Another modification might be to build outcome measures specific to the study in an attempt to obtain a better correspondence between program objectives and tested outcomes. These modifications, however, could be implemented only if the number of classrooms were reduced, since such revisions would increase costs significantly. We have not included these alternatives in our design because we feel that achieving an adequate number of classrooms is of prime importance, that program representativeness rather than geographic representativeness is essential, and that the commercially available outcome measures that we are recommending are superior to any measures that could be developed in the limited time available.

There is an additional alternative that would be a major departure from our basic approach. A consideration of its features will help to further clarify our own design. That alternative requires an operational definition of individualization leading to a way of categorizing programs as either individualized or not based on program descriptions. A further requirement is a set of procedures for observing classrooms that have adopted individualized and standard programs in order to determine if they are well implemented. To do so would require measures of how well the programs operating in classrooms conform to the design specifications of the developers of those programs. Once a set of classrooms that are well implemented versions of both types of programs is established, the next task would be to contrast achievement results, presumably in a covariance design that somehow took initial differences in abilities into account.

The main problem with this approach is that it assumes that within the two sets of classrooms there is a homogeneous treatment, and that there is a distinct, unique difference between the two sets. This, of course, is not true because all individualized programs vary in how they individualize and how well they individualize what they do individualize. Similarly, all standard programs individualize in some ways to some degree. Gibbons (1971) put it this way: "Identifying programs as individualized conveys so little information about them that any teacher can on some grounds claim to be individualizing instruction" (p. 15). He also pointed out that "individualized programs vary in the elements of instruction they individualize and the degree of individualization in those elements" (p. 54). Although adding the criterion "well implemented" could reduce the heterogeneity within treatment, it would still be substantial unless only a single program were examined.

We believe that too little would be learned from such a gross contrast of individualized and standard classrooms and that much information would be lost by eliminating the implementation variation that does exist, since this variation can be capitalized on to shed further light on which variations are effective and which are not in achieving student learning. The design proposed in this report will provide information both on the degree to which individualization is effective in improving the learning of all children, particularly those from homes that are disadvantaged in some respect, and on the specific ways in which it is most effective.

Section 2

Defining and Assessing Program and Support Variables

We recommend that both individualized and standardized classrooms be studied using a common set of variables. These variables, which are sensitive to both program design differences and variation in program implementation, will provide information on the processes actually taking place in the classrooms under investigation and the influence of these processes on student learning. The set of variables that we propose includes two types of observable classroom variables: program and support. Program variables are related to individualization; support variables are related to classroom practices that are not unique to individualization. Support variables in this field study serve the same function as control variables in an experimental design in that their inclusion in the study will make it possible to reduce the probability of alternative explanations for observed program effects.

The designer of any field research must anticipate alternative explanations and develop measures of the variables that they suggest. Then, through appropriate data analysis techniques and a convincing model of the phenomena under investigation, it is necessary to partition those effects that are attributable to the program variables from those that are due to the support variables and recognize which portion of the effects is confounded because of correlations among program variables and support variables. The analytical procedures for doing so are outlined in Section 6.

The framework for the study of classroom processes that we are proposing is provided by the model illustrated in Figure 2.1 (Cooley & Leinhardt, 1975; Cooley & Lohnes, in press). The purpose of this model is to explain the variation in student performance that occurs among classrooms following an extended period of instruction in those classrooms. The model specifies that criterion performance is a function of initial

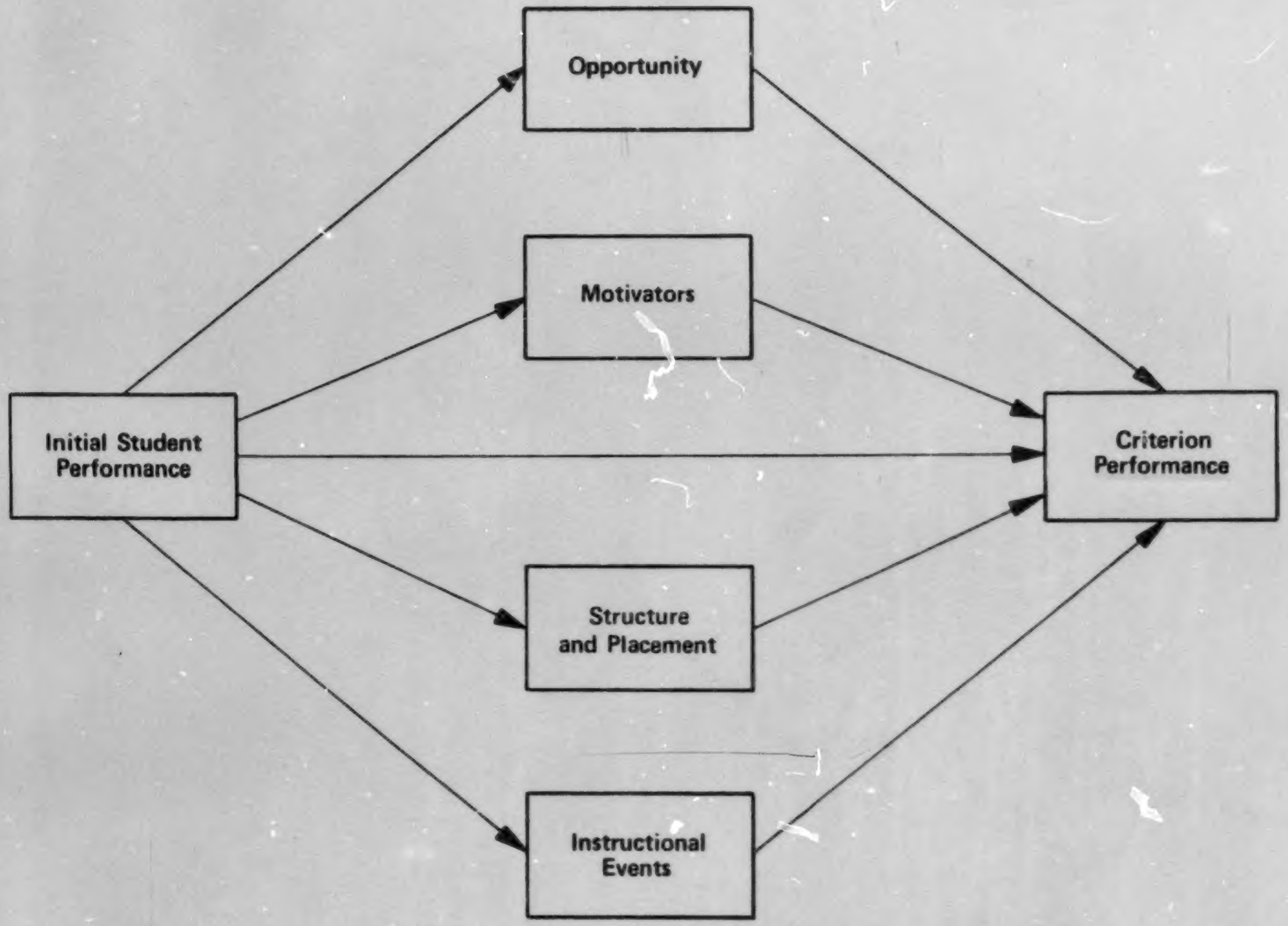

Opportunity
Motivators
Initial Student Performance
Criterion Performance
Structure and Placement
Instructional Events

Figure 2.1. Model of Classroom Processes

student performance and of certain classroom processes that occur in the interval between the assessment of initial performance and the assessment of criterion performance. Classroom processes are represented by four constructs that are assumed to affect the criterion performance: structure and placement, instructional events, opportunity, and motivators.

In terms of the model of classroom processes, program variables relate primarily to the structure and placement construct and to the instructional events construct. The support variables are reflected by the opportunity and motivators constructs. The 14 program and support variables that are included under the four classroom process constructs are described below, along with the measures and data collection techniques that can be used for their assessment.

Three data collection techniques are used to assess the process variables: interviews of the teacher, analysis of the curricula by curriculum experts, and videotaping of classroom activities. All three techniques have been developed and utilized over the past five years in extensive research on the impact of classroom processes on achievement (Leinhardt, 1975) and in detailed analyses of curricula with respect to the degree to which they incorporate principles of instructional theory (Holland, 1975).

Teacher interviews are extremely useful in determining specific classroom practices that are followed. In general, teachers attempt to provide accurate information, particularly if they do not feel threatened in any way by the questions asked and if they are aware that follow-ups will be made to ensure the validity of the information obtained. The fact that the interviews take place in the classroom also encourages teachers to be precise in their responses.

The purpose of curriculum analysis is to provide detailed information about the structure and quality of the curriculum materials. It is extremely important that the actual materials being used in participating classrooms be reviewed to see if they incorporate mechanisms for supporting individualization. In many cases, the form is present but the substance is lacking. Curriculum analysis also helps to cross-validate the information gathered from teachers and provides fine-grained information about the specific value of curriculum-designed instructional strategies.

Videotaping not only contributes to the cross-validation of teacher interview data but also provides unique information about classroom practices, particularly the quality of the instructional events. Taping requires fewer highly trained observers than in-class observation, eliminates the possibility of confounding observers with sites, and provides a permanent record of activities that will make it possible to monitor coding accuracy, recode ambiguous results, and reanalyze data at a later time using a different coding scheme.

There are distinct advantages to using this three-pronged procedure to gain information about classroom processes. First, information is gathered in such a way as to permit careful, reliable analysis in a different location. Second, by using three techniques that overlap, much of the information collected through one method can be verified by at least one other method. Finally, permanent records are created that will be available for reanalysis using a different statistical approach and asking different questions of the data.

A plan for the collection of data is presented in Section 5. The instruments themselves and directions for obtaining and coding the information required are presented in Appendix A ("In-Class Interviewing"), Appendix B-1 ("Analyzing Curricula"), Appendix B-2 ("Behavioral

Analysis of Curricula"), and Appendix C ("Videotaping"). All of the instruments have been designed or substantially modified for inclusion in the Individualized Instruction Study. The specific questions and assignments that are designed to collect data on the 14 process and support variables may be revised further by the contractor. However, our design requires that data be gathered, by one method or another, on all of the 14 variables.

Structure and Placement

Table 2.1 indicates the variables and measures for the structure and placement construct, along with the type of instrumentation that can be used to collect information on each measure. As this table indicates, the information is obtained through interviews and curriculum analysis. Videotaping is needed only to confirm interview data regarding the groups in which instruction is given.

The four program variables that are included under the structure and placement construct describe the structure, organization, and sequencing of the instructional materials, and the procedures for placement of the student in the curriculum. These variables are concerned with four major questions. First, is the content of the educational sequence specified? Measures of the "specification of objectives" variable that attempt to answer this question involve: the clarity of the curricular objectives; the specificity of the objectives (are they specific and content oriented, or are they general program goals); the frequency of presentation of new objectives; and the degree to which the teaching materials match the objectives. Second, are students matched to the curriculum according to their abilities and interests? Some of the "matching of students and curriculum" measures related to this question concern: the presence of placement, monitoring, and mastery assessment procedures in the curriculum; the frequency of monitoring; and the range of the frequency of monitoring. The third

Table 2.1

Variables, Measures, and Instrumentation for Structure and Placement

Variables	Measures	Instrumentation* Interview	Instrumentation* Curriculum Analysis	Instrumentation* Videotaping
Specification of objectives	Specificity of objectives	25, 32	7, 15	
	Clarity of objectives	26, 27, 33, 34	8, 9, 16, 17	
	Frequency of presentation of new objectives	30, 37		
	Degree to which materials match objectives	28, 35	10, 18	
Matching of students and curriculum	Presence of placement, monitoring, and mastery assessment procedures in curriculum		11, 19	
	Presence of placement procedures for use by teacher	29, 36		
	Presence in curriculum of method for assessment of initial abilities	39, 48	23	
	Presence of informal assessment procedure	41, 50		
	Presence in curriculum of method for assessing mastery	43, 52	24	
	Presence of curriculum-suggested alternative routes when difficulties encountered		27, 28	
	Presence of teacher-suggested alternative routes when difficulties encountered	59, 60, 61		
	Percentage of unique assignments	68, 69		
	Average number of days since last test	68, 69		
	Range of days since last test	68, 69		
	Average percentage correct on last test	68, 69		
	Degree of match between student and curriculum		12, 20	
	Overlap between placement test and curriculum		13, 21	
	Overlap between mastery test and curriculum		14, 22	
Sequencing and pacing of instruction	Clarity of sequence	45, 54	25, 26	
	Extent to which teacher follows sequence	46, 55		
	Extent to which teacher creates supplementary materials for improved sequencing	47, 56		
	Who makes sequencing decisions	57, 58		
	Presence of self-pacing	62, 63		
	Range of units completed	64, 65		
	Type of sequencing in the curriculum		29, 30	
	Amount of review	31, 38		
Grouping	Type of grouping for instruction	66, 67		
	Number of groups for instruction	66, 67		
	Number in groups for instruction	66, 67		
	Frequency of regrouping	66, 67		

---

*Numbers in each column refer to item numbers on the In-Class Teacher Interview Questionnaire in Appendix A and the Curriculum Analysis Questionnaire in Appendix B-1.

question with which structure and placement deals concerns mechanisms for making decisions with respect to the sequencing and pacing of instruction. Examples of measures for the "sequencing and pacing of instruction" variable are: the clarity of the sequencing of the curricular materials, the person who makes sequencing decisions, the presence of self-pacing, and the range of the learning rates of children in the classroom. The fourth question concerns the group in which children most often receive instruction (e.g., entire class, group of four individuals). Measures of the "grouping" variable concern the type and size of groups and the frequency of regrouping.

All four of the structure and placement variables relate to the individualization of instruction. That is, the more specific the objectives, the clearer the mechanisms for matching students and curriculum, the more opportunities for variation in sequencing and pacing based on student interest and learning needs, and the more individualized the instructional setting, the more individualized the learning process will be.

For specific examples of how these variables and their measures relate to individualization, consider the "matching of students and curriculum" variable. There are fourteen measures for this variable, seven of which deal with the presence of specific procedures in the curriculum or procedures constructed by the teacher that are designed to promote individualization by permitting easy matching of the student and the curriculum. One of these measures is "presence of placement, monitoring, and mastery assessment procedures in curriculum." If, for instance, a diagnostic mechanism exists in the curriculum, or the teacher has his/her own informal assessment procedure, then it is more likely that individual students' learning needs will be consistently and accurately assessed than if no such mechanism were present. If individual learning needs are measured, then there is a greater chance that these needs will be met than there would be if no testing were done. Once it has been established that a placement test does exist in the curriculum, information must be gathered on the overlap between the test and the curriculum. If the test does not reflect the curriculum content, then the chance of it being of value in meeting student needs is very low. But even the presence of effective matching procedures is not sufficient to guarantee individualization. Additional measures, such as "percentage of unique assignments," "average number of days since last test," and "range of days since last test" are needed to determine actual teacher practices related to these procedures. The more unique assignments made on any given day, for example, the greater the chance that students' needs are being met on an individual basis.

Instructional Events

Table 2.2 summarizes the variables and measures for the instructional events construct. Variables that are included under this construct describe the specific mechanisms, procedures, and actions that place the learner in contact with the feedback needed to progress toward a specified competency. The variables are of two kinds: interpersonal and curricular. The five interpersonal variables ("management information," "cognitive teaching to individuals or small groups," "cognitive teaching to whole class," "indirect teacher behavior," and "quality of teaching techniques") describe the quality of interaction between the teacher and the student. Interpersonal variables mainly concern the content, affect, and clarity of the teacher's instruction and the degree to which the teacher involves the student in active learning. Our assumption is that the greater the frequency of appropriate cognitive exchanges between student and teacher, the greater the chance that the student's learning needs are being met.

Table 2.2

Variables, Measures, and Instrumentation for Instructional Events

		Instrumentation*		
Variables	Measures	Interview	Curriculum Analysis	Videotaping
Management information	Frequency of management statements			3, 4
	Frequency of cognitive management statements			5, 6
Cognitive teaching to individuals or small groups	Frequency of cognitive statements alone			8, 9
	Frequency of cognitive questions			11, 12
	Frequency of child-initiated responses			17
	Frequency of child responses			14, 15
	Ratio of cognitive/management statements			24
Cognitive teaching to whole class	Frequency of cognitive statements alone			10
	Frequency of cognitive questions			13
	Frequency of child responses			16
Indirect teacher behavior	Frequency of silences or non-verbal time			18
	Frequency of personal statements			7
	Frequency of teacher errors			21
	Frequency of extended tutoring time			19
	Frequency of lecture time			20
Quality of teaching techniques	Degree to which active responses sought			26
	Degree to which teacher models responses			27
	Degree to which teacher refers to earlier curricular information			28
	Degree to which teacher focuses child's attention			29
	Degree to which teacher solicits child's opinion on correctness of response			30
	Degree to which teacher refers to earlier success			31
	Degree to which teacher tutors effectively			34
	Degree to which teacher manages class effectively			35
	Degree to which child receives feedback			36
	Degree to which teacher interacts with individual students			37
Quality of materials	Total number of appropriately determined responses per child		23	
	Ratio of appropriately determined responses to total		24	
	Consequence ratio		8	
	Predictive validity ratio		17	
	Discriminability ratio		12	
	Lowest of three ratios above		18	

---

*Numbers in each column refer to item numbers on the Behavioral Analysis of Curricula Data Sheet in Appendix B-2 and on the Videotape Summary Sheet in Appendix C.

The curricular variable ("quality of materials") relates to the quality of the instructional materials used. This variable deals primarily with the degree to which the assessment procedure is efficient with respect to the amount of time that the process detains the student from instruction, the degree to which it accurately matches the student to the curriculum, and the degree to which the materials elicit responses from students that are directly related to material that the student should learn.

The instructional events construct includes program variables rather than support variables because the vast majority of individualized curricula are designed on the assumption that teaching sequence and learning activities are closely matched to each learner's needs and interests. Although many individualized programs do not require students to be taught on a one-to-one basis, they do assume that students' needs and interests are being assessed on an individual basis and that instruction is prescribed based on that assessment. Many standardized curricula include aspects of individualized instruction, such as the administration of placement tests and posttests, but they do not serve the function of matching individual students to the curricular materials.

As an example of how the instructional events variables and measures relate to the individualization of instruction, consider the "quality of materials" variable and its measures. These measures ("total number of appropriately determined responses per child," "ratio of appropriately determined responses to total," "consequence ratio," "predictive validity ratio," and "discriminability ratio") assume that instructional materials support learning by arranging conditions under which the emission and reinforcement of the to-be-learned behavior are highly probable. Generally, in teaching, some material is presented to the learner; the learner interacts with the material in some covert fashion and then makes an overt public response that can be evaluated. Obviously, instructional materials differ in the extent to which overt student behavior is called for. When instructional conditions are appropriate, a higher per-child average of overt responses is expected to result in higher student achievement.

If the student's public response is incorrect, the desired private behavior has not occurred and thus has not been learned. Materials that generate many errors are not effective teaching materials. Effective materials are characterized by a low error rate. However, a low error rate alone does not indicate that the to-be-learned behavior has occurred. Effective teaching materials ensure that the learner is correct for the right reason. Problems in response contingencies are very common in teaching materials. Consider this anecdote by John Holt in How Children Fail (1964):

> Later, in music class, the children were asked to touch their toes when the teacher played a C. The teacher then played a little march, to which the children walked around. Every time she came to a C, she held it. Naturally, they touched them if any note other than C was held, and when C was played without being held, they ignored it. And this woman thought she was teaching them C! (p. 145)

In this case, the teacher thought the basis for the students' "correct" responses was the discrimination of a particular sound wave frequency, although she was also providing another basis for such correct responding, a held note. Yet her objective was not to teach the discrimination of held notes! A similar contingency problem occurs in many beginning reading programs that introduce new vocabulary words accompanied by pictures that the children easily use to guess rather than read the word.

The "total number of appropriately determined responses per child" is the total number of response opportunities available minus the opportunities likely to result in errors and the opportunities likely to be achieved by inappropriate behavior. It is expected that the per-child number of appropriately determined response opportunities will correlate positively with achievement. Yet it is possible that two classrooms with the same raw number of appropriately determined response opportunities per child could provide very different learning environments. Classroom A may have 12 per-child response opportunities, of which 10 are appropriately determined. Classroom B may have 40 per-child response opportunities, of which only 10 are appropriately determined. We would expect the 30 poorly designed response opportunities in Classroom B to have an adverse effect on the children's overall performance. Thus, in addition to a raw number, a ratio of appropriately determined responses to total response opportunities is included as a measure of the "quality of materials" variable.

These measures of appropriately determined response opportunities may be used to evaluate all teaching materials in terms of well-established principles frequently associated with programmed instruction: (1) individual teaching items should evoke the desired, to-be-learned behavior, and (2) the student should be able to give the required performance. However, a certain portion of the newer individualized or adaptive materials is not devoted to teaching, but rather to diagnostic testing. These materials contain diagnostic tests that either allow the student to skip upcoming teaching material or direct the student to specific needed material. Diagnostic test items have a very different function than do teaching items and, thus, require different criteria for their evaluation. It should be apparent that, if material is to be truly individualized, the decisions about what material a student is to receive or skip must be accurately made. Regardless of the structure of the material or how much testing and prescribing is done within a curriculum, that material is not adjusting to individual differences if the diagnostic decisions are not valid. The "consequence ratio," predictive validity ratio," and "discriminability ratio" are measures of how well the individualizing features of a curriculum route the student to needed teaching material and away from unneeded material.

Opportunity

Table 2. 3 lists the variables and measures for the opportunity construct. All three types of instrumentation are used to collect data on this construct, although teacher interviews are the primary source of information. Opportunity consists primarily of two variables: (1) the amount of time a student has to learn the subject matter, and (2) the degree of overlap between what is measured by the criterion test and what is taught in the curriculum. The assumption is that all other things being equal, a student who has a greater opportunity to learn material assessed by the criterion will know more of that material. For the purposes of the proposed study, the opportunity variables are regarded as support variables, since classrooms can be expected to vary with respect to the amount of time allocated to instruction and the amount of overlap between the curriculum and the criterion independent of the degree of individualization in the classroom.

There are 11 measures for the "amount of time" variable. It would seem that the amount of time that a student spends in reading and in mathematics could be measured directly and that only these two measures would be required for this variable. However, our experience indicates that it is extremely difficult to obtain accurate information on time spent in reading and in mathematics and that total time spent in learning is a more meaningful estimate of opportunity. Although we are including reading and mathematics times as measures because of their intuitive appeal and because our experience with these measures may not be

Table 2.3

Variables, Measures, and Instrumentation for Opportunity

Variables	Measures	Instrumentation*		
		Interview	Curriculum Analysis	Videotaping
Amount of time available to learn subject matter	Number of students enrolled	1		
	Average percentage of children present on two selected days	2		
	Average yearly attendance of students	3		
	Number of minutes in school day	8		
	Number of minutes in subjects	9, 10		
	Number of days in school year	4		
	Number of students transferring in and out	5, 6		
	Percentage of students on task			25
	Number of adults in the room	7		
	Percentage of new students contacted per time unit			22, 23
	Amount of homework assigned	11, 12		
Curricular overlap	Overlap of math materials with criterion test	71	1	
	Overlap of reading materials with criterion test	70	2	

---

[*] Numbers in each column refer to item numbers on the In-Class Teacher Interview Questionnaire in Appendix A, the Curriculum Analysis Questionnaire in Appendix B-1, and the Videotape Summary Sheet in Appendix C.

generalizable to the Individualized Instruction Study, we are also recommending that total time spent in learning-related activities be estimated from ten other measures, which concern the time spent in school, the percentage of time spent attending to what is being taught, the number of classmates a student must compete with in order to gain the teacher's assistance, and the amount of homework assigned.

The second opportunity variable, "curricular overlap," represents an attempt to answer the question, "To what extent is the information on the criterion measure taught?" Even with a superior teacher and curriculum, if what is tested is not taught, then the child will not have an opportunity to learn that material, at least in school. Curricular overlap is estimated in two ways. First, curriculum experts are asked to estimate which items on the criterion test are covered by each chapter or unit in each text or program used in the classroom. Then the amount of material covered by the class is collected and averaged (for example, out of 25 chapters in the math text, Group I covered 10, Group II covered 15, and Group III covered 22) and the overlap estimated. Second, teachers are asked to indicate those items on the criterion test that they feel were taught. This additional estimate of curricular overlap will provide information on test material that was taught even though it was not in the curriculum, and on material in the curriculum that was skipped by the teacher.

Motivators

The motivators construct consists of two sets of support variables: (1) variables related to aspects of the curriculum that support and encourage student learning, and (2) variables related to those aspects of teacher behavior (other than explicit instruction) and other interpersonal behavior that support learning. While there has been a great deal of research on motivation, there has been a limited amount of research on validating exactly which elements in a classroom motivate students to learn. Ideally, the study would measure only those variables that are known to have a dramatic impact on students' learning. But, given the limited amount of time to conduct the study and the limited amount of knowledge about effective motivators, we must take our best guess as to which elements in the environment are most supportive of student learning.

As indicated in Table 2. 4, information regarding aspects of the curriculum that encourage learning (e. g. , variation of format, speed of correction on product) is obtained through teacher interviews and curriculum analysis. Interviews and videotapes are used to collect information on interpersonal motivators, such as the degree to which the teacher uses praise and the degree to which he/she encourages the use of games and contests, self-evaluation and self-management, and peer tutoring. Preliminary data indicate that peer tutoring is a positive motivator. It may also be influencing the academic achievement of students because peer tutors are more effective teachers. However, since analysis of the specific technique of peer tutoring is not within the scope of the proposed study, the presence or absence of peer tutoring will be regarded only as an interpersonal motivator. The final interpersonal motivator is the amount of negative teacher behavior. This measure is negatively related to achievement, but it is still a motivator--a negative one. By negative behavior, we mean punishing behavior as distinct from correction, which need not be presented in a negative way.

Relating Process Constructs to Individualized Instruction

It may be useful at this point to relate our four process constructs (structure and placement, instructional events, opportunity, and motivators) to the objective of the Individualized Instruction Study--to investigate

Table 2.4

Variables, Measures, and Instrumentation for Motivators

Variables	Measures	Instrumentation*		
		Interview	Curriculum Analysis	Videotaping
Curriculum motivators that support learning	Degree of interest of materials	18, 19	3, 4	
	Degree of variation of format within mode	20, 21	5, 6	
	Number of modes of instruction	22, 23		
	Speed of correction on product	24		
Interpersonal motivators that support learning	Degree of use of peer tutoring	13		
	Degree of use of games and contests	14		
	Degree of use of self-evaluation	15		
	Degree of use of self-management	16, 17		
	Degree to which teacher uses praise			32, 33
	Frequency of negative teacher behavior			1, 2

---

* Numbers in each column refer to item numbers on the In-Class Teacher Interview Questionnaire in Appendix A, the Curriculum Analysis Questionnaire in Appendix B-1, and the Videotape Summary Sheet in Appendix C.

the effectiveness of individualized instruction. Our approach recognizes that there is considerable variation in how classrooms will be individualized, and that even so-called standardized classrooms will be individualized in some ways to some degree, depending upon the teacher and the program being used. Although such heterogeneity in treatment invalidates ANOVA-type experimental designs, our design capitalizes on this process variation by converting it into an opportunity to establish both the effectiveness of individualized instruction, however it may be defined, and the specific individualizing features that are most effective in promoting student learning.

In the RFP for the present project, NIE defined an individualized instructional program as one that includes:

1. Stated performance objectives.
2. Individual diagnosis of instructional needs, and instruction based on that diagnosis.
3. Progress tests that assess mastery at regular intervals.
4. Rate of progress and remedial instruction based on testing results.

The extent to which classrooms are operating programs with these features will be measured by the variables included in our structure and placement construct. If, in the data analysis, these variables are shown to be effective in explaining gains in reading and mathematics achievement, then individualized instruction, as defined by the RFP, is effective.

A behavioral analyst might define individualized instruction as instruction that "places the learner in contact with the reinforcement or feedback contingencies that the learner needs to progress toward a designated competency" (James Holland, personal communication). The extent to which the learner is exposed to the necessary contingencies is reflected in two variables in the instructional events construct ("quality of teaching techniques" and "quality of materials"). If these variables show a significant achievement effect, then individualized instruction, as defined by the behavioral analyst, is effective.

In Gibbons' (1971) descriptive analysis of individualized instruction, he began by indicating that the term "individualized" suggests "a specific instructional procedure adapted to the learning style and personality of each student" (p. vii). His analysis of the variety of ways in which programs and schools attempt to achieve this adaptation to individual students led to the specification of 15 dimensions along which they differ. He then plotted profiles for programs (e. g., Individually Prescribed Instruction, Winnetka Plan) and schools (e. g., Summerhill, Leicestershire) on these dimensions. Most of his dimensions are highly related to one or more of our process variables, and they sample all four of our process constructs. Some examples are: "time structure" (opportunity), "pace at which the materials are to be studied" (structure and placement), "activity that accompanies or follows study" (motivators), and "teaching method" (instructional events). Thus, to determine whether or not individualized instruction, as defined by Gibbons, is effective, one would assess the effects of our process variables that come closest to his dimensions of individualization.

Our own definition of individualized instruction is operationalized in terms of the program variables, which are reflected by the structure and placement, and instructional events constructs. That is, we consider classrooms to be individualized to the extent to which they include specific objectives, clear mechanisms for matching the student and the curriculum, and individual sequencing and pacing, and the degree to which the instructional materials and teacher are responsive to individual learner needs. To test individualization as we define it would require assessment of these variables.

One advantage of the design that we propose is that, as illustrated in Section 6, it allows for testing of the effectiveness of a variety of definitions of individualized instruction. Such tests are possible because we show how to dimension the relevant classroom processes, rather than just comparing classrooms that fit an arbitrary definition of individualized instruction with classrooms that do not. In addition, we recognize that it will be important to identify the kinds of available programs that seem to facilitate effective individualization. Therefore, another feature of our design, also illustrated in Section 6, is that it permits the identification of effective existing programs, as well as suggests how to improve programs by revealing the most effective mechanisms for individualizing instruction.

Section 3

Assessing Outcomes

The outcome of primary interest in the Individualized Instruction Study will be achievement in reading and mathematics; however, other outcomes must also be considered for inclusion in the study. This section describes the outcome measure that is recommended for assessing student achievement and the rationale for its selection. In addition, measurement of program effects on noncognitive student outcomes and on the total classroom environment is discussed.

Cognitive Outcomes

The design that we are proposing for the Individualized Instruction Study requires early fall and late spring testing. In addition, the number of school days between spring testing and fall testing must be stable across all participating classrooms in order to ensure the same number of potential days of schooling. It is also essential that test results be comparable. In order to stabilize the time of testing, to ensure comparability of results, and to minimize the possible effects of frequent retesting, local testing should be suspended in the classrooms included in the study and a single pre-selected achievement battery administered.

To adopt the tests and testing schedules of participating school districts appears to be a sensible approach from the standpoint of saving time and money as well as ensuring that local testing programs proceed uninterrupted. But in order to have complete and comparable data, the contractor who follows this approach would have to patch onto existing programs by gathering information on initial abilities or outcomes or both. Districts administer tests at a variety of times throughout the entire academic year, and they use a wide range of homemade and commercially available instruments. Some have no testing programs at all.

Use of an anchoring procedure that would permit the administration of several different batteries was also considered and found to be an unacceptable alternative. To patch together data from a variety of achievement test batteries requires a compelling statistical argument demonstrating the comparability of such data. To a certain extent, the Anchor Test Study (Bianchini & Loret, 1974) has provided a means whereby reading scores for the upper elementary grades on each of seven widely used achievement batteries can be equated. However, the results of this study are not useful for the present investigation. Data for the Anchor Test Study were obtained from and are applicable to children in grades 4 through 6. The population that we recommend for the Individualized Instruction Study is grades 1 and 3. Also, the Anchor Test Study concerned only reading achievement. The proposed effort is aimed at assessing achievement in both reading and mathematics.

Although a transformation system such as the Anchor Test could be incorporated into the present overall design, the expense of such an undertaking excludes it from serious consideration. Further, there is obviously greater confidence in the comparability of outcome data for any study when the instrumentation across all groups is as standardized as possible. The use of a single measure minimizes concern about variations in test quality, construction, reliability, validity, and administration.

Another consideration was the possibility of building and using criterion-referenced tests in the study. The argument for doing so is that achievement testing is valid only to the extent that the test is suited to a particular curriculum and accurately reflects the curriculum content. The argument has force depending upon the use that is being made of achievement test results. As diagnostic instruments, for example, criterion-referenced tests generally provide more useful information to a teacher than do standardized achievement tests. For the purposes of the proposed study, however, this consideration is not relevant. Given the fact that a variety of curricula, individualized and not, will be included in the study, one choice would be to construct a criterion-referenced test for each curriculum (an expensive endeavor requiring many months of effort), or, if the curricula have their own criterion-referenced tests, they could be used. This approach, however, would make comparison of results impossible.

An alternative might be to construct one criterion-referenced test for all curricula included in the study. This procedure would also be too time consuming. Test development could not begin before May 1976 and must be completed before the beginning of the 1976-77 school year. To design, pilot test, revise, produce, and disseminate a test in only four months would be impossible. Even if more time were available, the outcome measure that would be created would not necessarily be an improvement over, and perhaps not be as good as, standardized achievement batteries. For these reasons, criterion-referenced testing was rejected for use in the proposed study and the search for an acceptable standardized achievement battery undertaken. The search was begun by examining seven widely used achievement test batteries.

Two of the seven test batteries, the Iowa Tests of Basic Skills (ITBS) and the Sequential Tests of Educational Progress (STEP), were eliminated immediately because they did not include forms for the primary grades. The remaining five achievement batteries had levels and norm data appropriate for grades 1 through 3.

1. Metropolitan Achievement Test (MAT).
2. California Achievement Test (CAT).
3. Stanford Achievement Test (SAT).

4. Comprehensive Tests of Basic Skills (CTBS).

5. SRA Achievement Test.

Detailed reviews were made of the five batteries; these reviews are presented in Appendix D.

In evaluating each achievement test battery, several key points were considered. Among the most important was content validity. The five test batteries considered all make claims to high content validity, a claim based chiefly on how the tests were constructed. Tests that sampled a wide variety of curricula in use in widely divergent geographic areas of the United States and that were approved by curriculum experts and teacher educators were, in general, deemed valid tests.

Further support for the validity of a battery was found in data from analyses of item difficulty and discrimination and from subtest intercorrelations, and in other data obtained from the development of a standardization edition. Along with validity, the reported reliability estimates of each test battery were examined. The norm group statistics and data were also important considerations of the reviewers. Adequate sample sizes, acceptable methods of stratification, and efforts aimed at eliminating ethnic and racial bias were considered essential for an acceptable battery.

Surprisingly, format quality was quite variable among the five leading test batteries. Apart from considerations of the quality of print, illustrations, paper, and layout, the batteries were examined to determine the degree of sophistication in test-taking skills required of children taking the tests. Those tests having poor format quality or confusing layout, or requiring test-taking skills that were considered too sophisticated for young children were critically noted.

The feasibility of using a particular battery was another consideration. That is, it was necessary that those batteries being considered have available age-appropriate levels and alternate forms for those levels. Where alternate forms were not available, adequate demonstration of the feasibility of using different levels was required (e.g., Level A for assessment of first graders' initial abilities and Level B for assessment of their criterion performance). To use different levels, correlations between subtests of these levels must be high.

A final important focus of the test evaluations was the technical manuals that accompany the batteries. These manuals often fail to substantiate their claims with data. Therefore, full and complete reporting of data from validation studies on the instrument was deemed essential for attesting to the credibility of the instrument. Time, cost, and ease of administration were also taken into account and reported; however, standardized achievement test batteries are, in general, comparable in these areas.

On the basis of this evaluation, the Comprehensive Tests of Basic Skills is recommended for use in the proposed study. (A content analysis of this battery is presented in Appendix B-1.) The CTBS stands above other batteries as far as validity, reliability, and format are concerned. The development of the CTBS has been carefully documented, and all claims for the battery are supported by data reported in its technical manual. This manual is a model document and includes more information than is ordinarily found in a publisher's document. In item selection, sampling, and norming, special efforts were made by the CTBS developers to eliminate test bias against black and Spanish-speaking minority groups. Finally, items reflect a balanced sampling from the various curricular models current in American schools. Such curriculum sampling is unavoidable in the construction of a national standardized achievement test. Specific items, in fact, do test curriculum-specific learning skills (e.g., "phonics" versus "look-say" word-attack skills). Nonetheless, the representative sampling from various curricular models serves to place no student at a relative disadvantage simply because of his/her participation in a particular curriculum.

Noncognitive Outcomes and Classroom Environment

Although the RFP calls for consideration of the "nonachievement factors which contribute to classroom environment," it does not spell out what these factors might be. It was suggested that the designer review this area and propose what definitions and instrumentation, if any, should be included in the Individualized Instruction Study. Our approach to this task has been two-pronged: (1) to determine whether noncognitive student outcomes can and should be measured, and (2) to determine whether it is possible and desirable to assess the effect of programs on the total classroom environment.

We do not recommend that noncognitive student outcomes be assessed in the study for two reasons. First, although schooling, individualized or not, may indeed have an effect on some noncognitive outcomes, the theoretical basis for such a belief is not well developed. Without a sound basis, it is futile to attempt to measure noncognitive or social outcomes since it is not clear what to measure or how to make causal arguments if effects are found. A second argument against the testing of social outcomes is that their measurement in the primary grades is still in a primitive state.

Our consideration of noncognitive or social outcomes began with the generation of a list of outcomes that designers of instructional programs have claimed will be affected by their programs (e.g., self-concept, inquiry skills, autonomy). The next step was to locate instruments that purport to measure these specific outcomes. The short duration of the study ruled out the possibility of developing such instruments from scratch. Existing instruments were located, screened, and eliminated from further consideration if they failed to meet any one of the following criteria:

1. The instrument could not be highly correlated with reading and mathematics ability. If it were, it would measure little not already measured by the achievement test battery.
2. The instrument had to measure the social variables in question, i.e., it had to be valid as measured by standard measures of validity.
3. The instrument had to be reliable as measured by standard measures of reliability.
4. The instrument must have been designed or adapted for use in the primary grades.
5. The instrument must be usable from an administrative standpoint. This criterion would rule out instruments that are described in the literature but are otherwise untraceable, those that require an exorbitant amount of pupil/examiner time (in excess of three hours per pupil), and those that require a highly trained examiner or coder. A number of projective tests like doll-play were eliminated under this criterion.

The results of the search for an instrument that would meet these criteria were disappointing. Not one instrument of the many considered was totally acceptable. Table 3.1 lists some of the tests that were rejected and a criterion they failed. They may have failed other criteria, but this information was not recorded because the test reviewers eliminated an instrument upon failure to meet one criterion.

Table 3.1

Instruments Considered and Criterion They Failed

Instrument	Criterion Failed*
Early School Personality Questionnaire	2
Carel Instruments, Form S-B	3
Edwards Personality Inventory	4
The Thomas Self-Concept Values Test	3
California Test of Personality, Primary Form AA	3
How I See Myself, Elementary Form	4
Problem Expression Scale	4
Sears Self-Concept Inventory	2
Attitude Toward School, School Sentiment Index, Primary	3
Coopersmith Self-Esteem	4
Piers-Harris Children's Self-Concept	4
Driscoll Playkit	5
Animal Crackers	2

---

*Numbers refer to the criteria listed on the previous page.

One of the tests considered, though it failed the criterion regarding content validity, did include several worthwhile features. This test, "Animal Crackers," is reviewed in detail in Appendix D, but mention will be made here of several of its characteristics. First, in its five scales, it covers a range of social outcomes from Purposiveness to School Enjoyment and, therefore, casts a sufficiently wide net as to pick up any positive or unintended negative effects of school. Second, the test items, in the main, address school-centered behaviors and attitudes so that they are likely to tap effects related to school rather than home or community effects. Third, the instrument can be administered in both the fall and spring, thereby making statistical control possible. Finally, the test's predecessor, Gumpgookies, has been widely used in previous evaluations of compensatory education programs. Although we are not recommending that Animal Crackers be included in the study, the contractor may be interested in reviewing it to get a sense of the limitations of even the "best" of the currently available measures of noncognitive student outcomes.

Our recommendation regarding assessment of the effects of schooling on the total classroom environment is that no attempt be made to measure these effects. There are two reasons for this recommendation. First, what these effects might be is not clear, so it would not be possible to construct a priori measuring instruments to tap them. Instead, these effects would have to be determined in situ. To do so would require one highly trained observer/ethnographer per instructional model. These individuals would have to identify, describe, and then measure the effects of schooling on the environment. Further, their efforts would have to be very well coordinated to ensure comparability of results. The cost of such an endeavor would be prohibitive, even in $1.5 million study.

A second and more compelling reason for not studying the unspecified environmental effects of schooling is logical rather than logistical.

Instructional programs are designed to impact academic achievement. Changing the classroom environment is the means to accomplish this end. We cannot treat environmental effects as both independent variables (means) and dependent variables (ends). We must assume that if a given instructional strategy has any impact on the classroom environment, then this impact will be reflected in the cognitive outcome measures.

We do recognize, however, that value can be attached to both means and ends. Everyone "knows" that the ends do not justify the means. In fact, most parents are as concerned about the kinds of experiences their children have in school as they are about what their children learn. We believe that the way to respond to this need is to use our process variables as descriptors of what classrooms are like and as predictors of how much is learned by children in those classrooms.

Section 4

Sampling and Securing the Cooperation of Schools

Sampling

Based on both statistical and cost considerations, it is recommended that 400 classrooms be included in the Individualized Instruction Study. This sample size is large enough to enable the contractor to include 16 variables (initial abilities, criterion performance, and the 14 program and support variables) in the data analyses. It is small enough to make intensive examination of classroom processes affordable in a $1.5 million study.

Further, it is recommended that 200 of these classrooms be at the first-grade level and 200 at the third-grade level. Grade 1 is recommended because it is very important to assess the effects of educational programs as early as possible. Not only is grade 1 the most critical year for beginning reading--a most important and highly valued skill, but it is also the first grade level at which reasonable confidence can be placed in test results. In addition, it is the earliest grade for which there is reasonable consensus as to the objectives of the reading and mathematics programs.

Given that grade 1 was to be included, the choice was then whether to include other grade levels and, if so, which one(s). It did not seem necessary to include every grade suggested for consideration in the RFP, i.e., kindergarten through grade 4. Studying more classrooms at a few grade levels seemed more desirable than studying several levels with fewer classrooms per level. Including at least one other grade level in addition to grade 1 was deemed valuable because it is possible that effective classroom processes may be different at different grade levels.

Grade 3 was selected as the other grade level to study. It was chosen over grade 2 because it was felt that there is considerable advantage in spreading the information over the age ranges (grade levels) of interest. Further, some individualization processes become more fully operational at grade 3, where the need for intensive small-group instruction is diminished. Also, in grade 3, achievement expectations are different than they are in the earlier grades. Reading tests, for example, change from the first to the third grade, making it important to examine changes in effective classroom practices as desired reading outcomes change.

Grade 3 was preferred to grade 4 for two reasons. First, by grade 4 many programs require that children move from classroom to classroom or teacher to teacher for different subjects. As a result, it is difficult to assess the relative independent effects of classroom processes for given children. Second, the entering abilities of higher-grade students become increasingly confounded with previous schooling, thus overwhelming the potential for explanation of end-of-year performance by current processes.

Kindergarten was not selected for investigation because the initial ability testing of kindergarten children is extremely difficult, particularly since group testing is impossible. In addition, the vastly different expectations as to what is to be accomplished in kindergarten make it difficult to arrive at a sensible measure of kindergarten outcomes.

In defining a plan for selecting specific classrooms, there were several major considerations:

1. The primary ways in which classroom processes can vary must be represented in the classrooms selected, since the effects of classroom process on achievement cannot be established unless there is adequate variation in process.

2. The number of curricula in use in the sample classrooms must be kept to a manageable size so that it will be possible to: (a) estimate the degree of correspondence between objectives covered by the curriculum and objectives sampled by the outcome measures, and (b) estimate the instructional efficiency of the materials used. These analyses would be very expensive and time consuming if a different curriculum were operating in each participating classroom.
3. Geographic representativeness is not essential, since the Individualized Instruction Study is not aimed at estimating national parameters such as the frequency with which various forms of individualized instruction are used in compensatory education.
4. The classrooms to be selected should include children eligible for compensatory education programs. However, it is not necessary or even desirable to restrict the sample to compensatory programs. Low achieving children will be present in almost any classroom. The study does not require representativeness of present compensatory programs, but it does require representation from a variety of approaches to classroom instruction.

Lists of specific schools that would participate in the Individualized Instruction Study are not being provided at this time for two reasons: (1) the identity of the study contractor is unknown, and it is recommended that participating schools be located in the immediate vicinity of the contractor's headquarters; and (2) it did not seem appropriate to seek school cooperation for a study that is only in the design stage. What is provided in this section is a set of procedures for identifying schools that could be asked to participate in the study and for securing their cooperation. It should be noted at this point that sampling for this type of study is necessarily an interactive process, and that continual adjustments must be made between what is desirable and what is possible.

The first step in identifying possible schools is to determine which two or three states in the immediate vicinity of the study contractor (or data collection subcontractor) to include. To do this, the states that show adequate variation in program and type of poverty (e.g., urban, small town, rural) must be identified. If, for example, the study contractor were located in Pennsylvania, one might begin by considering Pennsylvania (two major cities and several smaller cities, small town and rural settings) and West Virginia (small cities and towns and rural settings including parts of Appalachia). If upon further examination, an adequate variety of programs cannot be identified in these two states, then Ohio, for example, could be added. Even if one state shows adequate variation, at least two states should be sampled so as to make the study results more convincing.

The next step is to identify school districts in the states selected that appear to be operating individualized programs. To do so, the following lists should be obtained: (1) schools participating in the national Follow Through program; (2) schools implementing one of the nationally available individualized programs (e.g., Individually Prescribed Instruction, Individually Guided Education, Project PLAN); and (3) schools participating in District Survey I.* These lists and descriptions of the programs are available from the United States Office of Education, commercial publishers, and NIE, respectively. From the lists and descriptions, it will be possible to determine the number of school districts that could participate in the study, the features of individualization that appear to be incorporated in the programs being implemented in those districts, and the types of programs in operation (i. e., reading and/or mathematics). Programs descriptions often do not reflect the exact nature of the program. However, a general notion of the program's characteristics can be obtained from this source.

---

*Students who will participate in the instructional setting contrast described in Section 6 will also be selected from District Survey I schools.

Table 4. 1 illustrates the results of our review of program participants and descriptions for Pennsylvania, West Virginia, and Ohio. The programs listed in this table are all operating in Pennsylvania and/or Ohio. The information that we reviewed indicated that only one individualized program is operating in West Virginia, so that state was excluded as a possible study site. Eight programs were identified that incorporate at least one of five features of individualization that can be estimated from program descriptions: (1) stated performance objectives that are cognitively oriented, (2) systematic diagnosis of the instructional needs of individual children and provision of instruction based on that diagnosis, (3) systematic monitoring procedures to assess mastery at regular intervals, (4) procedures for the adjustment of instruction with respect to rate and/or content based on systematic assessment of student progress, and (5) specific instructional materials that have been developed or identified for use in individualized instruction. These five individualizing features are all associated with the structure and placement characteristics of a program. Since ensuring variation in classroom processes is the primary objective of the sampling plan and since these features represent the major ways in which individualized programs vary, it is fortunate that adequate variation in the five features can be established early in the sampling process. The other constructs in the classroom processes model

Table 4.1

Programs Operating in Ohio and Pennsylvania School Districts

Program	Objectives	Diagnosis	Monitoring	Adjustment	Specific Materials	Reading/ Math	Number of School Districts
Behavioral Analysis Approach (Kansas)	X	X	X	X	X	R/M	1
Developmental-Interaction Approach (Bank Street)		X			X	R	1
Englemann/ Becker Model (Oregon)	X		X	X	X	R/M	1
Individualized Early Learning Program (LRDC)	X	X	X	X	X	R/M	28
Individually Guided Education (Wisconsin)	X	X	X	X	X	R/M	56
Mathemagenic Activities Program (Georgia)	X		X		X	M	1
Project PLAN (Westinghouse)	X	X	X	X		R/M	1
Responsive Educational Program (Far West)				X		R/M	1

that we are recommending (i.e., opportunity, motivators, and instructional events) cannot be estimated from program descriptions. This is not a problem, however, since our experience has shown that variation on these three constructs can be found in any set of 200 classrooms.

As Table 4.1 indicates, there are at least 90 school districts in Pennsylvania and Ohio that could be asked to participate in the study. Ninety districts will not be needed, but not all districts will agree to participate. Replacement can be achieved by selecting alternative districts that are operating similar instructional programs and that maintain variation in type of setting (urban, small town, rural). We recognize that there may be attrition of schools during the course of the study. However, the sample size that we are recommending will permit up to 20 percent attrition.

Once identified, candidate school districts should be contacted to determine: (1) their willingness to participate in the study; (2) the number of first- and third-grade classrooms in the district that are operating an individualized program in reading and/or math; (3) the individualizing features of that program; and (4) the number and type of standardized programs operating in grades 1 and 3 in the district (i.e., programs exhibiting none of the individualized features listed earlier). From the school districts that are willing to cooperate in the study, 200 first-grade and 200 third-grade classrooms must be selected. Table 4.2 illustrates the procedure for sampling first-grade classrooms within a school district. The general recommendation is to select up to 20 classrooms per grade level per district, about half of which are operating individualized programs, and to keep the number of districts to under 20. In selecting individualized classrooms, preference should be given to those classrooms that exhibit a unique combination of individualizing features in their program, that include instruction in both reading and math,

Table 4.2

Illustration of Sampling First-Grade Classrooms Within Districts

School District	Individualizing Features	Reading/ Math	Urban/ Small Town/ Rural	Individualized (I) Classrooms Available	Standardized (S) Classrooms Available	Classrooms Sampled I	Classrooms Sampled S
A	1 2 3 4 5	R/M	U	12	50	10	10
B	1 2 3 4 5	R	S	9	20	3	6
C	1 2 3 4 5	R	R	6	8	6	6
D	1 2 3 4	R/M	S	15	0	0	0
E	1 3 5	R/M	R	7	0	7	0
F	2	R/M	U	14	45	10	10
G	4	M	U	18	36	10	10
H	1 3 4 5	R/M	U	20	60	10	10
I	1 3	R/M	S	15	10	10	10
J	1 2 3 4 5	M	S	20	0	0	0
K	1 2 3 4 5	R/M	U	25	90	10	10
L	1	R/M	U	18	55	10	10
M	1 2 3 4 5	R/M	R	4	8	4	8
N	1 2 3 4	R/M	U	15	72	10	10
						100	100

and that are located in districts where standardized classrooms are available for comparison purposes. Also, there must be adequate representation of type of setting. At least 50 percent of the classrooms should be located in urban settings.

After specific individualized classrooms are identified, school achievement test means should be obtained for the schools that include these classrooms, as well as for schools in the district that are operating standardized programs. Standardized classrooms should be selected from schools with test means as similar as possible to those of the individualized classrooms.

Securing the Cooperation of Schools

Once the states that will be included in the study have been identified, the Project Director and/or School Coordinator and, if possible, a representative of NIE should personally contact the chief state school officers to enlist their cooperation. Meetings are a far more effective means of obtaining cooperation than are written or telephone requests. It should be emphasized that the purpose of the study is not to evaluate schools or classrooms, but rather to identify effective mechanisms for improving instruction. At the local level, initial contacts should be made with district superintendents and school principals by the Project Director and/or School Coordinator. In addition, there are a number of incentives that would increase the probability of a district agreeing to participate in the Individualized Instruction Study, as well as ensure cooperation throughout the course of the study. Not all school personnel will be interested in the same incentive; therefore, it will be important to match incentive to the individuals involved. Some possible incentives are as follows:

1. Make available data summaries considered to be important by the district. This information would be available within the data collected for the study.
2. Offer to share with the cooperating district the study results as soon as possible. Delayed reports are not well received.
3. Make training sessions required for school personnel as attractive as possible, e.g., schedule at their convenience, compensate for time if the sessions are not held during working hours, conduct in comfortable facilities.
4. Provide copies of previously prepared research reports or papers that are relevant to the study being conducted.
5. Extend invitations to attend, as guests of the contractor, a symposium that would focus on the nature, purpose, and importance of the study, and would include participation by NIE representatives.

Other possible incentives may be identified in discussions with school personnel.

Once schools have agreed to participate, the next step is to obtain the cooperation of participating and nonparticipating teachers in each school. Both groups of teachers should be informed of study activities. In addition, since extra demands will be placed on participating teachers by the testing, interviews, and videotaping in their classroom, it is suggested that they be compensated in some way. Our recommendation is that teachers be paid an honorarium of $50 each upon completion of the spring testing. Doing so will cost $20,000, but it could mean the difference between a successful or unsuccessful $1.5 million effort.

Section 5

Collecting Data

In the study that we are proposing, data will be collected through in-class interviewing of teachers, analyzing curricula, videotaping classroom activities, and administering the Comprehensive Tests of Basic Skills. This section presents recommendations regarding the scheduling of all data collection activities, the personnel required, the training of these personnel, and procedures for monitoring the quality of the data collection effort.

In-Class Interviewing

In-class interviewing of teachers will provide data on three of the process constructs: structure and placement, opportunity, and motivators. The In-Class Teacher Interview Questionnaire is included in Appendix A, along with directions for obtaining and coding the information required. Interviewing will take place twice during the school year: October-November 1976 and March 1977. Site Coordinators, under the supervision of the Field Coordinator, will be responsible for scheduling interviews with school personnel at each site. (The positions of Site Coordinator and Field Coordinator are discussed in Section 7, along with other key project positions.)

Ideally, the interviewers should be persons who have a good idea of what goes on in the classroom, e.g., former teachers. The Field Coordinator is responsible for the training of all interviewers. He/she should conduct one or two training sessions to review the questionnaire and the directions for obtaining and coding the information. It may be beneficial for the interviewers to practice administering the questionnaire several times to each other. If possible, arrangements should be made for the interviewers to administer the questionnaire to teachers in schools that are not participating in the study prior to interviewing participating teachers. It should be made clear to the interviewers that they should probe for accurate and complete information rather than adopting the traditional survey approach of trying to fit the first response given into a predefined category.

An interview will take approximately 1.5 hours per classroom to complete. That is, to collect data from 400 teachers will require 600 hours in the fall and in the spring. Using the same personnel to conduct both interviews will be an advantage in terms of ease of training and confidence in the reliability of the information obtained.

Coders of the interview data should be college students or clerks. Depending on contractor preference, the interviewers may also serve as coders. The Field Coordinator is responsible for the training of all coders. The coordinator (and perhaps the interviewers) should explain the coding of each question to the coders in the fall. In addition, it may be useful for him/her to meet with them in the spring to explain the various methods for combining fall and spring codes into a single code. Any questionnaire answers that are unclear to the coder should be clarified with the interviewer who completed the questionnaire. Coding a single questionnaire requires approximately one hour. That is, to code questionnaires from 400 classrooms will require 400 hours in the fall and in the spring. Using the same personnel for coding both fall and spring data is recommended.

As the interviews are coded, they should be grouped into batches of ten. In order to maintain high levels of quality control, one out of each batch of interviews should be recoded. If errors are discovered, the entire batch of interviews coded by the particular individual making the detected errors should be recoded. All coders should initial their work so as to permit the identification of individuals who make consistent errors and, if necessary, the reassignment of tasks. In addition, it is recommended that informal lists of item responses be compiled by batch so that obvious errors can be detected before the data source and the data become too separated.

Analyzing Curricula

Information on all four of the process constructs will be provided through an analysis by curriculum experts of the curricula in use in participating classrooms. A Curriculum Analysis Questionnaire is included in Appendix B-1, along with directions for obtaining and coding the information required. Additional information will be gathered through a behavioral analysis of each curriculum. Instrumentation and directions for conducting such an analysis are presented in Appendix B-2.

Preparations for curriculum analysis must begin before the end of the 1975-76 school year. That is, in May 1976 a list must be compiled of the reading and mathematics curricula that are expected to be used in each classroom so that the materials needed by the curriculum experts can be ordered. In the fall, a final list of curricula should be prepared and any materials ordered that were not on the previous list. It is recommended that curriculum analysis be initiated in September 1976 and completed in December 1976.

The curriculum experts should have experience in instructional design. Training could take one to three weeks, depending on its intensiveness and the background of the experts. It is suggested that the Field Coordinator, who is responsible for training, become familiar with a text entitled The Analysis of Behavior in Planning Instruction (Holland, Solomon, Doran, & Frezza, in press). This text is in the form of self-instructional units, several of which might serve as the basis for training on the behavioral analysis of a curriculum (e.g., Units 18, 19, 20, 28, 31). These units could be supplemented with a minimum of didactic material and a great deal of elementary math and reading lesson samples on which the curriculum experts would be asked to count response opportunities, identify response contingencies, and estimate potential error rate. Behavioral analysis training could also be provided by a number of individuals or their students (e.g., Eva Baker, Center for the Study of Evaluation; James Holland, University of Pittsburgh; Susan Markle, University of Illinois at Chicago Circle).

The more familiarity the curriculum expert has with the curricula used in participating classrooms, the more rapidly he/she will be able to evaluate them. It is estimated that an expert will require 10 to 20 hours to complete and code the Curriculum Analysis Questionnaire for each classroom for each curriculum (i.e., for reading and for mathematics). An expert who is well acquainted with the curriculum will probably require 10 hours, whereas an expert who has never seen the curriculum before may require as much as 20 hours. A behavioral analysis of the curriculum will require 2 to 4 hours when all children in a classroom have identical assignments. When all children have different assignments, the analysis may take 10 to 20 hours.

Quality control procedures similar to those recommended for the coding of interview data should be followed. However, because of the greater amount of time required to analyze a curriculum, only one of every twenty-five analyses should be redone. If errors are discovered, five more analyses should be reexamined. If additional errors are found, the entire batch should be analyzed again.

Videotaping

Although videotaping of in-classroom activities will yield information on or confirm previously collected data on all four of the process constructs, its main purpose is to clarify the instructional events in a classroom. Videotaping will occur both in the fall (October-November 1976) and in the spring (March 1977). In addition, some videotaping should be done in nonparticipating classrooms before the end of the 1975-76 school year to provide practice for the videotaping team as well as tapes that can be used in training sessions for the observers. The Site Coordinators will be responsible for scheduling videotaping at each site.

The collection of data through videotapes requires videotapers, assistants, and observers. The tapers should be knowledgeable in the use and repair of audio-visual equipment. The assistants should have experience working in schools as teachers or field researchers. Videotape observers or coders should be experienced in in-classroom observation. It is strongly recommended that the same observers, as well as videotapers and assistants, participate in data collection in both the fall and spring. To tape one classroom will require 4 hours in the fall and in the spring (i.e., 2 hours for reading and 2 hours for mathematics). Coding a videotape from one classroom will take 3 hours in the fall and in the spring, or a total of 6 hours during the school year.

In Appendix C are directions for producing videotapes, for obtaining measures from videotapes, for coding videotapes, and for maintaining quality control. The information presented is sufficiently detailed to provide the basis for training sessions to be conducted under the supervision of the Field Coordinator.

Administering the CTBS

As indicated in Section 3, the CTBS is recommended for use in measuring students' initial abilities and end-of-year achievement. The proposed testing schedule is as follows:

	September Pretest	May Posttest
Grade 1	CTBS-Level A Form S	CTBS-Level B Form S
Grade 3	CTBS-Level 1 Form S	CTBS-Level 1 Form T

It is suggested that two subtests of the grade 1 battery not be administered. These subtests are "Language" in Level A and "Language II" in Level B.

Although using classroom teachers as testers is admittedly the most convenient and inexpensive way to implement the testing program, this approach also has the most potential for introducing bias into a testing program. For this reason, it is suggested that participating teachers not administer the test batteries to their class. Instead, outside testers should be hired. One possible source of testing personnel is the pool of available substitute and/or retired teachers in a locality. The Site Coordinators, under the supervision of the Testing Coordinator, should contact local school boards participating in the study and request a list of substitute and/or retired teachers who are available. (The position of Testing Coordinator is described in Section 7.) Test administrators can then be selected from this list.

In addition to having nonbiased testers, it is important that all test administrators be thoroughly trained in a standardized testing procedure. It is highly recommended that all testers be trained by the same individuals. To ensure that all testers will follow the same model, a film or videotape should also be produced to demonstrate model testing behavior.

An adequate training program and competent supervision of the test administrators by the Testing Coordinator and Site Coordinators should serve to establish a high level of quality control in test administration. As an added check, the Testing Coordinator may ask the Site

Coordinators to monitor testing procedures in a sample of classrooms. A checklist of administration could be drawn up based on the standardized procedure set forth in the training sessions. Any variation in this procedure would be noted and corrected by the Site Coordinators. However, given the expense of such a procedure and the type of training and supervision proposed, this extra check of test administration should be implemented at the contractor's discretion.

Another quality control procedure, one that is strongly recommended, is initial verification of the tests before they are returned to the Testing Coordinator. Answer booklets without names, stray marks, missing data, etc., should all be corrected before students are dismissed and the testing considered complete. The testers should be responsible for this phase of quality control.

Just as important as the quality of test administration is the quality of test scoring. At the present time, considerable debate surrounds the issue of hand versus machine scoring. Our experience, as well as the experience of other organizations such as the Stanford Research Institute, has indicated that there is considerable savings in cost, as well as greater accuracy, with hand scoring. The superiority of hand scoring has been found to be especially true in the scoring of younger children's protocols, grades 1 to 3, where answer sheets cannot be used. Since every test must be scored at least twice in order to validate scores, it is recommended that an initial hand scoring followed by a second hand-scored validation be adopted as the most economical and accurate procedure.

Workshops

It is suggested that a workshop be held for data collection personnel in July 1976 to familiarize them with all aspects of the field work.

A further recommendation is that district supervisors, principals, and perhaps Site Coordinators be invited to a workshop at the contractor's office to acquaint them with the purpose, scope, and time demands of the study. This workshop should take place in August 1976.

Section 6

Reducing and Analyzing Data

This section summarizes the procedures that we recommend for reducing classroom process measures to a small number of variables, as outlined in Section 2, and for analyzing the resulting variables to reveal the effectiveness of various classroom processes in the teaching of reading and mathematics. In addition, a small-scale study is described that should be conducted parallel to the main investigation of classroom processes. This special study contrasts in-classroom individualization strategies with the practice of "pulling" disadvantaged children out of classrooms for special tutoring, a frequently used Title I program approach.

Reducing Data

To describe and analyze the large amount of information that will be collected in the Individualized Instruction Study requires that the classroom data be systematically reduced to a manageable number of process dimensions. It is conceptually impossible and statistically undesirable to analyze 200 measures separately, for example. Fortunately, the measures described earlier were constructed with this consideration in mind. The basic approach that we propose is to reduce the measures along the lines of the six constructs of the classroom process model illustrated in Figure 2.1. The steps involved in data reduction are: elimination of unusable measures, preliminary correlation and partial correlation analysis by variables within constructs, inspection and reflection of measures, plotting and transformation of data, development of standard scores with unit variance, and combination of measures to form variables. By combining the data in a manner that preserves the initial "meaning" of the numbers, the final results will be more interpretable. Also, unpacking of the data for more detailed analysis of any particular construct can be carried out in a straightforward manner.

Elimination of measures. In any large-scale data collection effort, some of the information will be unusable for one reason or another. The purpose of this first step is to eliminate such measures. Means, standard deviations, and skewness should be examined, and measures with zero or near zero variance should be deleted. Measures with unusual skewness should be examined for outliers to check for obvious errors (e.g., if inspection reveals one classroom with 300 students, a check should be made since that value is highly unlikely for that measure). Another way of identifying outliers is to compute frequency distributions and scan them for "unusual" values. Obviously, not all unusual values are incorrect, but simple clerical errors can be detected in this way.

Correlation and partial correlation. The next step in data reduction is to inspect the correlations among the measures for each of the variables. Correlations should be computed on the measures along with data from initial and end-of-year student performance aggregated as classroom means. These correlations should be inspected, particularly with respect to unsually high relationships between initial abilities and process measures. Identifying such relationships will be helpful in the later interpretation of commonalities, should they exist. Also, the signs of the relationships between measures and outcome residuals need to be inspected. This task can be accomplished by partial correlation, removing initial ability from process and outcome, or by correlating process with achievement gain. The objective is to identify measures that do not conform to theory with respect to their effect on achievement. This procedure is a conservative form of criterion scaling, which at least ensures that a measure does not need to be reflected before it is combined with other measures. A negative partial would indicate the need to consider reflecting a measure.

Data transformation. One of the assumptions of the analytic approach that we are advocating is that each process measure relates to the outcome measure in a linear fashion. It is possible, however, that some of the measures relate in a nonlinear way. The purpose of this step is to determine whether any measures need to be transformed so that the eventual combination of measures into variables does not result in loss of information. Procedures for determining whether a measure needs to be transformed and for selecting appropriate transformations have been summarized by Tukey (1970) and need not be detailed here. After the measures are transformed to make them additive, they need to be standardized with a zero mean and unit variance prior to combining them.

Combination of measures. The recommended procedure for combining measures into variables is to simply add the measures for each variable after they have been adjusted to unit variance, with separate scalings for math process and reading process. This procedure reduces the classroom data from over 75 measures down to a manageable number of variables. These variables should be combined with initial abilities and outcomes for data analysis. Separate analyses should be conducted for reading and for mathematics, and for grade 1 and grade 3.

Principal components analysis was one of the alternatives considered for combining measures. The first principal component represents a most reliable single dimension of what is common to that set of measures. However, it does not incorporate variance in measures that are unrelated to the set. For example, obtaining a single dimension of socioeconomic status from a set of correlated indicators such as income, occupation, and education can be accomplished by using the first principal component for that set. If a measure in that set does not correlate with other measures in the set, it will not load on the first component. This is not serious if one is attempting to scale a construct like socioeconomic status, which is viewed as a single dimension that "explains" the relationships among the measures and/or that represents the variance common to the set.

Measures of a particular classroom variable are not necessarily correlated with each other. Thus, in combining classroom measures, what determines whether a measure should be combined with others in the set is whether or not the measure has face validity for the variable being assessed. In Table 2.2, for example, there are 10 measures to be combined into a "quality of teaching techniques" variable. Teachers high on one measure will not necessarily be high on others. If, however, one assumes that the more of those behaviors present in the classroom, the more effective the teaching, then a sum across all such measures would be an indication of "quality of teaching techniques."

Analyzing Data

The RFP requires a plan "for arguing causal relationships between program and outcome variables" in a nonmanipulatory survey of existing classroom practices, despite the difficulty of such arguments under the canons of statistics (as reviewed by Cronbach & Furby, 1970). Since everyone knows that correlation cannot prove causality, what is required are analyses that create the strongest valid presumption of causality. Although there are analyses of correlations that permit valid and stronger presumptions of causality under ideal research conditions (e.g., path analysis), the actual conditions for the Individualized Instruction Study call for a conservative method of analysis. On the technical side, this is because the measurements employed will possess only partial validities for the constructs they represent, and will have variable and unknowable degrees of unreliability. On the policy side, this is because there is less hazard to the nation when research underestimates for Congress the impact of particular arrangements for education than when research claims unreplicable effects. Commonality analysis provides an appropriately conservative approach to arguing causality in the results of the study.

Commonality analysis is a comparison of statistical models for the data. When it creates a presumption of causal efficacy for program variables, it does so by reducing the probabilities for alternative explanations of the observed educational development of pupils. The basic fact about survey research data is that the possible causes of outcomes as described by alternative models are confounded to some extent. Rather than attempt hazardous unconfounding by algebraic tricks, commonality analysis represents the confounded portions of effects on outcomes as separate partitions of the outcome variance (called "commonalities"). These confounded portions are not included in estimation of the irreducible effects of the separate possible causes (called their "uniquenesses"). The conservativeness of the method resides in this use of minimum rather than maximum estimates of separate effects. The advantage to the method rests with the visibility of the commonalities; other conservative approaches tend to mask them.

When the uniqueness for program variables has been computed from the data as a result of comparisons of models using the combinations of initial abilities, support, and program variables, the precise interpretation of it will be that the uniqueness of the program variables is the portion of the variance in educational development that can only be explained by employing the measurements of program variables in a complex model for the outcomes. No model employing only the initial abilities and support variables can account for this portion of the variance in educational development. Granting that this does not prove the degree of causality of the program variables, to the extent to which the models under comparison are judged complete and competent it justifies a presumption of degree of causality.

While the external validity of a survey study is dependent on the characteristics of the samples achieved, the internal validity is dependent on the qualities of the models compared. Causal arguments will be most convincing in the presence of complete and technically competent contrasted models. The omission of probable causes that are known to be correlated with the hypothesized cause of central interest is devastating. Unconvincing operationalizations of constructs or ridiculous unreliabilities are debilitating. Exhaustiveness and redundancy are the rules to follow in making observations. Parsimony and organization are achieved by combining detailed measures prior to commonality analysis. That is, constructs for the initial abilities, support, and program variables are scaled as linear functions of exhaustive redundant observation scales. The gain in reliability for a positive linear function of a set of positively correlated indicators is substantial and should not be overlooked.

The model that has guided our design of the Individualized Instruction Study postulates that a particular set of program variables describing ways and degrees in which instruction is individualized is essential to the explanation of pupil learning in reading and mathematics that occurs over the study year. The model holds that this set of program variables combines with two other sets of variables, the initial abilities and support variables, to generate pupil learning. The alternative models that have to be shown to be inadequate in the data analysis hold that: (1) initial abilities are sufficient explanation of pupil learning, and (2) abilities together with support variables are sufficient explanation of learning. Since these are simpler explanations, their truth probabilities can only be reduced by showing appreciably greater power for the complex model involving the addition of information about program variables.

Table 6.1 illustrates the most general result from a sequence of possible commonality analyses, in which the total variance explained (i.e., the squared multiple correlation) is .68, using gain in mathematics achievement as the dependent variable and 15 abilities and process variables as the predictors. (These variables are listed in Table 6.5.) Table 6.1 indicates that 18 percent of the variance in improved mathematics performance can be attributed to classroom processes. It also shows that 22 percent is due to process and/or ability, and is the result of correlations between process measures and initial ability that occurred among participating classrooms. This commonality is the extent of the in-common effect for process and ability that cannot be sorted out in this particular study. It is partly a function of the fact that better classroom processes tend to be available to more able students. One objective of the sampling design is to plan classroom process variation so that it is as independent of the initial abilities of students as possible.

A process effect of 18 percent compares quite favorably with the less then 5 percent that is currently claimed as the amount that can be attributed to school differences (Averch et al., 1972; Coleman et al., 1966). Should this study find a process effect that is larger than 5 percent, it would be an important finding for restoring confidence in the notion that what happens in school does make a difference in what children learn. It is more likely that such an effect will be found by directly measuring the instructional processes that are operating at the classroom level, rather than by measuring variables that characterize the school.

Tables 6.2 and 6.3 are hypothetical outcomes at the next level of generality. They are relevant to the question of whether the guiding model,

Table 6.1

Partition of Variance for Gain in Grade 1 Math Achievement*

(Illustrating unique contribution of classroom processes)

Source	Proportion of Variance
Initial Abilities	.28
Classroom Processes	.18
Commonality	.22
Error	.32
Total Criterion Variance	1.00

---

*This table, as well as the other tables in this section, presents the hypothetical results of an analysis of grade 1 mathematics data. Similar analyses should be conducted and reported on grade 3 mathematics data, and grade 1 and grade 3 reading data.

Table 6.2

Partition of Variance for Gain in Grade 1 Math Achievement

(Illustrating weak contribution of program variables)

Source	Proportion of Variance
Initial Abilities	.28
Support Variables	.15
Program Variables	.03
Commonality	.22
Error	.32
Total Criterion Variance	1.00

Table 6.3

Partition of Variance for Gain in Grade 1 Math Achievement

(Illustrating strong contribution of program variables)

Source	Proportion of Variance
Initial Abilities	.28
Support Variables	.08
Program Variables	.10
Commonality	.22
Error	.32
Total Criterion Variance	1.00

which asserts that program variables are essential to the explanation of pupil learning, is supported by the results. The results in Table 6.2 do not provide strong support for the model, since the contribution of program variables is small. The outcomes illustrated in Table 6.3, on the other hand, do seem to justify the model.

However, several considerations must influence what has to be a subjective decision regarding the usefulness of the program variables. Assuming that smaller increments of explanation are useful when larger proportions of total outcome variance are accounted for, whereas only larger increments of explanation can impress when the best model is weaker, then one must view the uniqueness for the program domain in the perspective of the total of all uniquenesses and commonalities, or the "total variance explained." Small improvement of a strong prediction system can be worth considerable cost. Only substantial improvement of an inherently weak prediction system may be worth the same cost. In addition, it is necessary to know which of the program variables seem to be most involved in producing the uniqueness for the program domain. Some program variables are more amenable to policy manipulation, or more economical to manipulate, than others. If the indicated manipulations are easy and cheap, they may be justified by a relatively small uniqueness. Also, it is important to know what values of initial abilities and support variables seem to combine with what specific values of program variables to produce the best results, as well as which combinations are to be avoided. It must be remembered that the best model for the data is likely to be an interactive one that does not hold out the same rewards for a given value of a program variable over all values in the ranges of support variables and initial abilities. This interaction can be represented in the model by scales fabricated as cross-products of other scales. Another consideration concerns which aspects of a complex criterion performance seem to be most and least influenced by the program variables.

In mathematics, for example, if simple computing is influenced but problem solving is not, policy makers may not value the learning that is influenced as much as they do the global concept of "mathematics." All this means that there must be other tables that "unpack" the summary of results in tables such as 6.2 or 6.3.

Table 6.4 reports hypothetical results in terms of the four process constructs and illustrates how the commonalities themselves can be of interest. All but the motivators construct seem to be important in explaining gain in mathematics achievement. This lack of a unique contribution for motivators may be the result of unreliable measures or of confounding with other predictors. Whether or not confounding exists can be determined by an examination of the commonalities. As the table indicates, the motivators construct (3) is confounded with initial abilities (1) and opportunity (2), as well as being part of an in-common effect for all five of the predictor sets. Thus, although motivators may be important in influencing achievement in mathematics, the correlation between motivators and the other predictors makes it impossible in these particular results to identify a unique effect for motivators.

Table 6.4 also illustrates the results of the analysis that would be performed if one defined individualization in terms of the structure and placement construct. The Table 6.3 analysis of the combined effect of the program variables represents the test of individualization that we propose, since we define individualization in terms of the program variables in both the structure and placement, and instructional events constructs. The analysis reported in Table 6.4 distinguishes between these two constructs; this hypothetical illustration shows that both constructs are important predictors of student achievement.

Table 6.5 illustrates the most detailed level for unpacking the summary statistics. Hypothetical uniquenesses are presented for the

Table 6.4

Partition of Variance for Gain in Grade 1 Math Achievement

(Illustrating contribution of the four process constructs)

Source	Proportion of Variance
1. Initial Abilities	.28
2. Opportunity	.08
3. Motivators	.00
4. Structure and Placement	.06
5. Instructional Events	.04
Commonalities	
1 and 2	.04
1 and 3	.03
1 and 5	.05
2 and 3	.02
4 and 5	.01
1, 2, 3, 4, and 5	.07
Error	.32
Total Criterion Variance	1.00

Table 6.5

Variables with Full Model* for Partition Reported in Tables 6.3 and 6.4

Initial Ability	Uniqueness
1. Classroom means	.27
2. Classroom variance	.01
Support Variables	
3. Amount of time available to learn subject matter	.03
4. Curricular overlap	.05
5. Curricular motivators	.00
6. Interpersonal motivators	.00
Program Variables	
7. Specification of objectives	.03
8. Matching of students and curriculum	.00
9. Sequencing and pacing of instruction	.03
10. Grouping	.00
11. Management information	.00
12. Cognitive teaching to individuals or small group	.01
13. Cognitive teaching to whole class	.00
14. Indirect teacher behavior	.00
15. Quality of teaching techniques	.03

---

*One of the process variables, "quality of materials," can only be measured on programs that include diagnostic tests and procedures for individual decision making. A separate analysis should be run with classrooms operating such programs. This analysis would make it possible to determine the variation in instructional effectiveness among materials produced for individualized programs and the effect of this variation on achievement gain.

15 predictors of classroom gains in mathematics learning for 200 first-grade classrooms. These uniquenesses indicate that certain variables seem to be irrelevant for this criterion, or at least make no independent contribution to its prediction. The uniqueness for initial abilities suggests that what a class holds for knowledge and ability at the beginning of the instructional year strongly prefigures what it is likely to gain during the year as new knowledge and abilities. The uniqueness for the two opportunity measures ("amount of time available to learn subject matter" and "curricular overlap") reaffirms the notion that pupils learn what they are taught to the extent to which time is devoted to that teacher-learning effort. The uniquenesses that describe the most payoff for the individualization of instruction are for "specification of objectives," "sequencing and pacing of instruction," and "quality of teaching techniques."

The computer program that we recommend be used for the commonality analyses (Veldman, 1975) computes all possible commonalities, but only reports those larger than .01, since the number of combinations becomes quite large as the number of variables increases. Commonalities smaller than .01 can also be quite important, however. It is possible that some variables can only occur in combination with others, and thus can only exhibit an in-common effect. These variables should not be dismissed as unimportant if a unique contribution is not found. For example, "matching of students and curriculum" (variable 8) shows no unique effect, but might well reveal a joint effect with "specification of objectives" (7) and "sequencing and pacing of instruction" (9). This joint effect would be indicated if high scores on (8) only occurred in combination with high scores on (7) and/or (9). Attention to regression coefficients alone would not reveal this type of effect.

In inferring the relative importance of the predictors from their uniquenesses, one must also consider the reliability with which the process variables were measured. It is possible that a variable may not appear to be important simply because it was less reliably measured. Some investigators have suggested the possibility of building reliability information into the estimation of regression coefficients (e.g., Wiley, 1973). We have suggested that the data be analyzed using commonality analysis, but, clearly, alternative analysis strategies should be explored by the contractor.

Variance is a bread-and-butter idea for educational researchers, and explanation of criterion variance is the essence of successful curriculum research for them. Persons in Congress cannot be expected to share this viewpoint. The abstract notion of explained variance can be grounded in real terms for policy makers by developing examples of predicted outcomes for various vectors of predictor values by means of the raw-score regression equation. For a constant level of initial abilities and support variables, various levels of program variables can be demonstrated as possible manipulation strategies. The disclaimer that policy manipulations are not likely to have quite as much effect as predicted from the regression equation has to be attached, but these demonstrations may provide the best advice to policy makers the study can afford. If interaction variables have been validated, the demonstrations will have to be made for several levels of the interacting variables.

An auxiliary to the model for classroom learning should be a model for the variability among classrooms in the program domain. This model (technically a partitioning of canonical redundancy) would explain from initial abilities and support variables some of the program variability. It would help policy makers to understand who gets what in school under present sociopolitical realities, and might illuminate policy options.

Persons in Congress no doubt will be most interested in the effects of program variables on the gain in achievement, as estimated from classroom means. Yet program variables may have socially significant effects on the shape of classroom learning distributions, as well as on the means. Of Fisher's four descriptors (Lohnes, 1972) of classroom learning distributions, the variance and skewness have been found to be the most useful in describing the shape of the distribution. Suitable graphic demonstration of the program effects might be useful. Computer plotting graphics will be needed to reveal the shape implications of various vectors of values for the raw regression equation. It would be remarkable indeed if the study were to document graphically that individualization, without segregation of low ability youths, tends to shorten the left tail of learning distributions in the manner required for "mastery learning." Elongation of right tails would indicate the nurturance of unusual talent. It will also be possible to explore whether or not instruction that adapts to individual differences in learners increases or decreases learner variability.

The Individualized Instruction Study is a study of what children learn in the course of a school year. It is vital that the criterion variables represent what children learn in the year of the study, rather than what they had learned in six to eight years of life prior to the study year. It is well known that strong models for initial abilities can be made. The only way to provide a criterion that represents what children learned in one year, rather than what they know from all their years of learning, is to subtract what they could do at the beginning of the year from what they could do at the end of the year. This change in performance is the natural criterion language for the study, and persons in Congress will appreciate it, even if statisticians sometimes decry it.* It should be noted that most of the criticism to which change scores have been subjected over the years has been applied to situations where the individual pupil is the unit of analysis. The change scores proposed for this study will be changes in classroom means. Presumably, the reliability of these change scores will be reasonably high. Whatever the technical characteristics, they represent what Congress wants to know about. Some risks have to be taken.

---

*Recently, however, this question has been reexamined, and we do not now feel quite so apologetic regarding the use of gain scores. See, for example, Overall (1975), Kenny (1975), and Richards (1975).

Program profiles. Another analysis that should be conducted as part of the main investigation of classroom processes involves a comparison of variables that prove to be effective in supporting learning. For example, taking the uniquenesses reported in Table 6.5 as an indication of the most effective process variables, values on variables 3, 4, 7, 9, and 15 could be averaged for all classrooms that are using a particular program. These results, when properly interpreted, could have important implications for schools that must decide which program to implement.

Figure 6.1 illustrates the profiles for three hypothetical programs (A, B, and C). The location and shape of the distributions for classrooms using these three programs are represented by schematic box plots (Tukey, 1972). The horizontal line cutting through each box represents the median of the classrooms operating that program, the height of the box represents the 25th and 75th centile points, and the vertical lines extending from the box indicate the range of observed values. The differences on "time," although important in explaining achievement, are not relevant to the selection of a program, since the amount of time allocated to reading and math can be manipulated independent of the instructional program that a school uses. The "overlap" measure needs to be considered at least to the extent required to understand unusually low scores. A low score would suggest that a program does not include many of the objectives included in the achievement test. If the omitted objectives were viewed as important, then a program that is low on "overlap" would not be an appropriate choice. Of the five effective process variables, "objectives" and

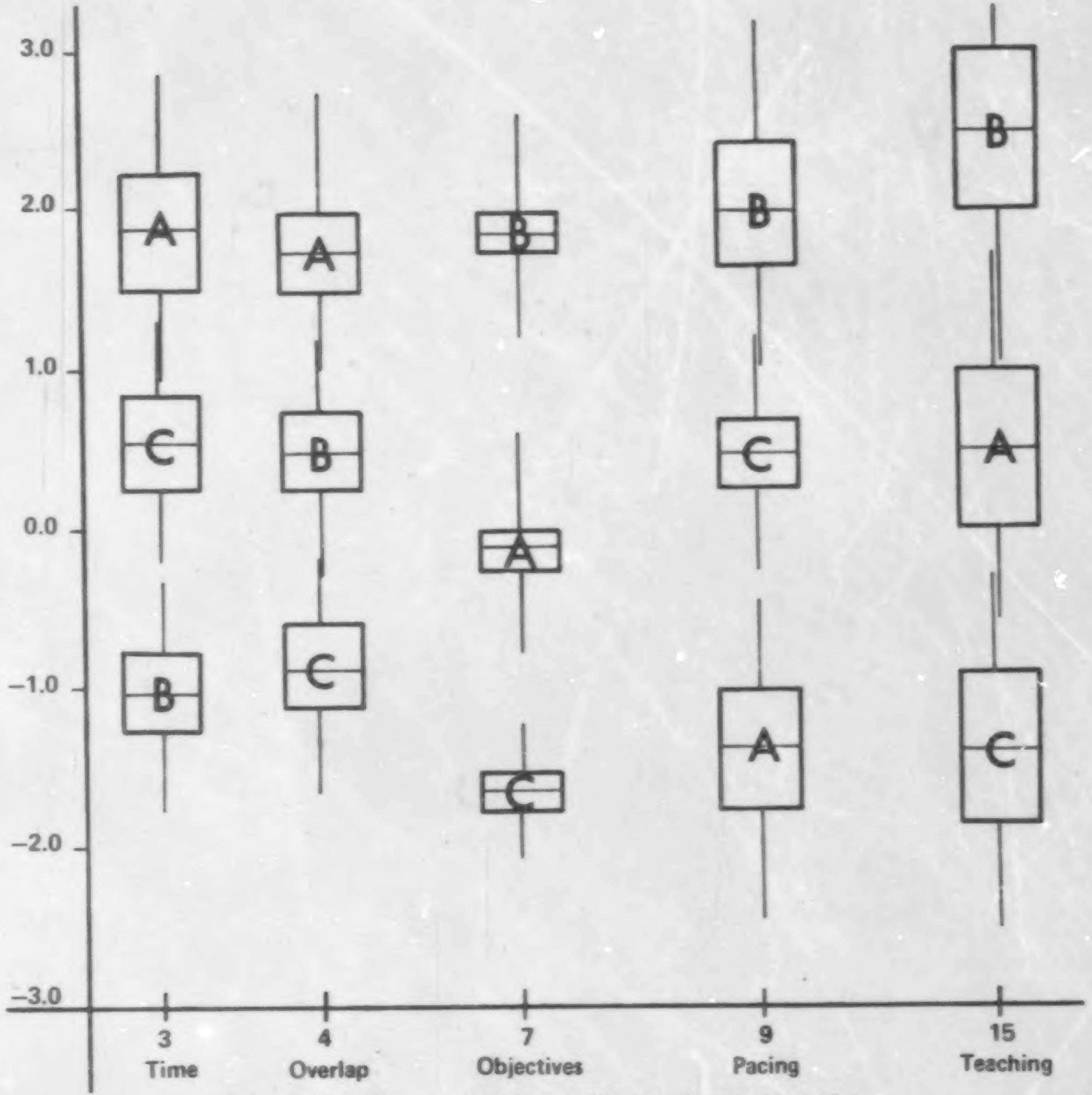


Figure 6.1. Program Profiles on Effective Process Variables

"pacing" are the ones that are most directly a function of the program being used, and thus are important in program selection. In this illustration, program B has the highest scores on these two important features. The fifth variable, "teaching," may be related to the kinds of teacher behavior that the program or its accompanying training package encourages (or allows), or it may be related only to the characteristics of individual teachers. Thus, this variable should not be given too much emphasis in selecting among available programs. A study that simply contrasted achievement gain for these three programs could not provide this kind of information. Program A, for example, might appear to be as effective as the others primarily because the teachers who happened to be using it devoted time to mathematics to the exclusion of instruction in other important skills. Although we do not expect that any single program incorporates all of the features of effective instruction, this type of program profile analysis could indicate which programs appear to be the best bets.

Instructional Setting Contrast

In addition to showing the extent to which classroom processes can affect children's learning and the relative effectiveness of different individualizing procedures, it is important for the Individualized Instruction Study to attempt to show that low ability children can perform as well in classrooms in which effective processes are operating as they perform in Title I tutorial situations. For this aspect of the study, the general approach is to identify tutorial programs in District Survey I that are in the same geographic region that is under investigation in the Individualized Instruction Study. From our examination of Title I files in Pennsylvania, we estimate that most communities have Title I programs that would be suitable for including in this contrast. Only 100 tutorial students in grade 1 need to be involved, since the student will be the unit of analysis. These students should be included in both the fall and spring testings. Tutorial programs will not, of course, receive intensive process observation, but they should be visited during the school year to make sure that they are consistent with available descriptions. It will be important to determine if tutorial work is instead of or in addition to regular classroom instruction in reading and/or mathematics. If possible, the tutorial participants should be selected from programs in which the tutorial work displaces classroom instruction in the same subject.

After the process data on classrooms in the Individualized Instruction Study are analyzed, the tutorial children should be matched on abilities with 100 first-grade students from classrooms that are revealed to be high on effective process measures. The contrast between the two groups of students could be a simple covariance design. The expectation is that low ability children will perform at least as well in a regular classroom as in a Title I tutorial program, particularly if that classroom was rated high on the process measures shown to be effective in the main analysis. It will also be possible, without collecting additional data, to contrast students in the tutorial mode with students from classrooms that exhibited varying degrees of effective process. This contrast would allow one to examine just how effective classroom process must be in order to be at least as effective as, if not better than, the very expensive tutorial mode.

Subsequent Analyses

Given the rich dataset that will be collected in the study that we are proposing, numerous analytical approaches are possible and should be tried after the main analysis is completed. What any good design must do is ensure that analytical schemes are available for answering the major questions of the study, but much more investigation will be possible and desirable after the main effects have been established. For example, it might be desirable to rerun the main analysis, including in the estimate of achievement gain only those students who would be considered eligible for compensatory programs (using a uniform criterion of eligibility across all classrooms) and adding the mean initial ability level of the entire classroom to the predictors. If the classroom ability level proved to have a positive impact on the achievement of eligibles, it would suggest the importance of heterogeneous grouping in facilitating the development of low achieving children. Interaction terms could also be tried. To do so requires introducing the product of two variables as an additional predictor. It is possible, for example, that heterogeneity is effective only in the presence of good structure and placement.

One might also conduct an analysis in which the program that each classroom is using is included as a dummy variate. If the resulting set of variates shows an important program effect in the presence of our process measures, then the conclusion would be that there are important differences among programs that are not being captured by the process measures. That finding would have important implications for program selection as well as indicate the need for revision of the classroom processes model.

Section 7

Project Organization, Staffing, and Schedule

Organization and Staffing

Any organizational structure that we might suggest for the Individualized Instruction Study will be and should be modified by the contractor to reflect the nature of the contracting organization and the competencies of the individuals who will be responsible for major functions. If any subcontractors participate in the study, other modifications may have to be made. For these reasons, only a very general project structure is presented (Figure 7.1). The necessary qualifications of each of the key individuals included in this structure are briefly described below, along with the functions that they will perform.

The Project Director must have extensive experience in the administration of large-scale research projects. In addition, this individual must have special competencies in the areas of field research in education, statistical analysis, evaluation methodology, and instructional design. The Project Director will have overall responsibility for the study, with direct responsibility for report preparation and coordination with NIE. Further, he or she may also be qualified to perform one of the other functions specified below.

There are at least three reasons why it is very important that an exceptionally well qualified individual direct the project. First, the Individualized Instruction Study involves the collection and analysis of enormous amounts of data, the management of a large number of project personnel, and coordination with many outside individuals and groups. Second, the study must be completed in only 18 months, which is a short period of time for such a mammoth undertaking. Third, and perhaps most important, the results of the study will likely impact on thousands of children across the nation. Thus, the study must be well executed

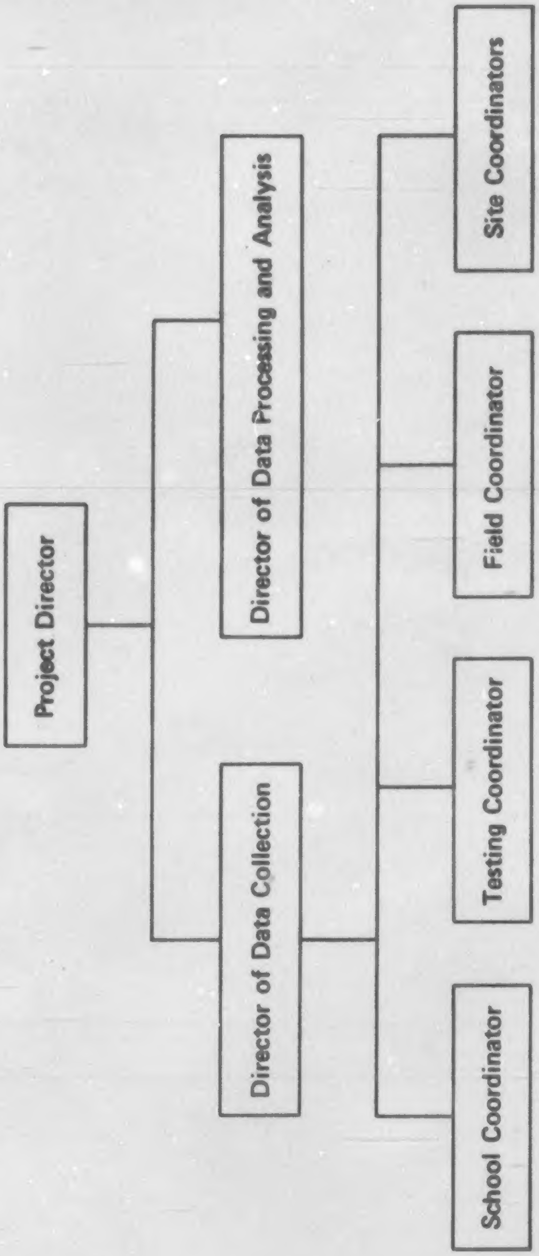

Project Director
Director of Data Collection
Director of Data Processing and Analysis
School Coordinator
Testing Coordinator
Field Coordinator
Site Coordinators

Figure 7.1. General Project Structure

and well reported. To achieve this requires direction by an extremely knowledgeable and experienced individual.

The Director of Data Collection must have previous experience in field research in education that included the administration of standardized tests and the conduct of classroom observations. He or she must be familiar with the types of problems that can occur in a large-scale data collection effort and with techniques for solving these problems. This individual's responsibilities include supervision of the School Coordinator, Testing Coordinator, Field Coordinator, and Site Coordinators, as well as the reporting and preliminary interpretation of classroom data.

The School Coordinator will serve as liaison between study personnel and participating school districts. He or she will be responsible for identifying and securing the cooperation of specific schools and for making all necessary arrangements with state and local school administrators for testing, videotaping, and interviewing to take place in the classrooms (e.g., obtaining school board permission). Because the School Coordinator must work closely with educational administrators, it is strongly recommended that this individual have experience working in a public school system, preferably as a district superintendent or principal of a large school.

The Testing Coordinator will be responsible for all activities related to the administration of achievement tests in the fall and spring, including ordering and disseminating tests, coordinating the hiring and training of test administrators, arranging for test scoring, and ensuring that adequate quality control procedures are followed in test administration and scoring. As indicated previously, the test administrators, possibly substitute teachers, should be hired specifically for this study. Approximately 100 administrators will be required at $500 per person.

The Field Coordinator will supervise all activities related to the collection of process data through interviews, curriculum analysis, and videotaping. Examples of these activities including disseminating all necessary data collection materials, arranging for the purchase and maintenance of videotaping equipment, obtaining curriculum materials for analysis, and coordinating the hiring and training of the field staff. This field staff will conduct in-class teacher interviews, analyze curricula in use in participating classrooms, videotape in-class activities, and code all process data. The number of individuals and person-months of effort will depend on the functions that the contractor assigns to each person (e.g., interview teachers, or interview teachers and also code the data), the time schedule adopted, the quality control procedures implemented, etc. Our estimate is that a field staff will be required of at least 5 full-time individuals for the duration of the study and 35 full- and part-time individuals for varying amounts of time up to 15 months.

The Site Coordinators will serve as liaisons between the project staff and school principals and teachers at each site. They will, for example, arrange the scheduling of in-class interviews and videotaping, monitor and assist in the collection of both process and achievement data, and respond to requests for information about the study from school personnel or parents. Approximately 40 Site Coordinators, possibly school-related personnel, will be required. It is anticipated that over the course of the study they will work the equivalent of 8 person-months each.

The Director of Data Processing and Analysis must have extensive experience in supervising the processing and analysis of large amounts of education-related data, with special competencies in statistical analysis and evaluation methodology. His or her staff will include a full-time programmer, a small group of coding clerks who will prepare data for keypunching, and a keypuncher. In most organizations, coding and keypunching are done by a support group. Thus, the coders need not be full-time members of the project staff.

Schedule

A proposed schedule of project activities and products to be delivered to NIE is presented in Table 7.1. In preparing this schedule, an attempt was made to specify products that would be the direct result of ongoing study activities (e.g., a confirmed list of participating schools, principals, and teachers), rather than products that would have to be prepared solely for the purpose of keeping NIE informed. Only two major reports are required--a status report in February 1977 and a final report in July 1977.

Table 7.1

Project Schedule

Date	Activity	Product to Be Delivered to NIE
February 1976	Complete sampling plan.	
March 1976	Prepare tentative list of schools, principals, and teachers.	Tentative list of participants
April 1976	Secure cooperation of participants.	Confirmed list of participants
May 1976	Prepare tentative list of curricula in each classroom. Order these materials.	Tentative list of curricula to be analyzed
	Begin making arrangements for fall data collection (e.g., staff, instruments, equipment, schedule, parental permission).	Data collection schedule
	Prepare practice videotapes.	
June 1976	If necessary, request permission from NIE to administer standardized test other than test previously agreed upon.	
July 1976	Conduct workshop for data collection personnel.	Agenda and summary of workshop presentations
	Place orders for tests.	
	Select tutorial sample for instructional setting contrast.	

Table 7. 1 (continued)

Date	Activity	Product to Be Delivered to NIE
August 1976	Conduct workshop for local site personnel at contractor's office.	Agenda and summary of workshop presentations, or workshop attendance by NIE representatives
	Complete arrangements for fall data collection.	
	Train test administrators and curriculum experts.	
September 1976	Obtain class rosters.	Class rosters
	Prepare final list of curricula in each classroom. Order any materials not previously ordered.	Final list of curricula to be analyzed
	Begin curriculum analysis and complete overlap estimates.	
	Administer achievement tests.	
	Schedule make-up tests.	
	Cross-check tests against class rosters.	Estimate of percentage of missing data by classroom
	Begin test scoring.	

Table 7.1 (continued)

Date	Activity	Product to Be Delivered to NIE
October 1976	Complete administration of make-up tests.	
	Complete test scoring.	
	Complete training of videotapers and interviewers.	
	Begin videotaping and interviewing.	
	Continue curriculum analysis.	
November 1976	Complete videotaping and interviewing (do not tape or interview two days before or after Thanksgiving recess).	
	Analyze fall test results.	Results of fall testing for both main study students and tutorial sample
	Train interview coders and begin coding of interview data.	
	Continue curriculum analysis.	
December 1976	Complete curriculum analysis.	Results of curriculum analysis
	Train videotape observers.	
	Begin coding of videotape data.	
	Continue coding of interview data.	

Table 7.1 (continued)

Date	Activity	Product to Be Delivered to NIE
January 1977	Complete coding of videotape and interview data.	
February 1977	Prepare status report and outline of final report.	Status report and outline of final report
	Complete schedule for second round of videotaping and interviewing.	
March 1977	Complete videotaping and interviewing (do not tape or interview two days before or after spring break).	
	Complete coding of videotape and interview data.	
	Complete scheduling for spring testing.	
April 1977	Construct final dataset minus outcomes.	
May 1-17, 1977	Administer achievement tests.	
May 18-31, 1977	Complete test scoring.	
June 1-15, 1977	Complete data analysis.	
	Select individualized sample for instructional setting contrast.	
June 16-July 31, 1977	Conduct instructional setting contrast.	
	Prepare final report.	Final report

References

Averch, H. A., Carroll, S. J., Donaldson, T. S., Kiesling, H. J., & Pincus, J. How effective is schooling? A critical review and synthesis of research findings. Santa Monica, Calif.: Rand Corporation, 1972.

Bianchini, J. C., & Loret, P. G. Anchor Test Study (Vol. 1). Berkeley, Calif.: Educational Testing Service, 1974.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. Equality of educational opportunity. Washington, D. C.: U. S. Government Printing Office, 1966.

Cooley, W. W., & Leinhardt, G. The application of a model for investigating classroom processes. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1975.

Cooley, W. W., & Lohnes, P. R. Evaluation research in education. New York: Irvington Publishers, in press.

Cronbach, L. E., & Furby, L. How should we measure "change"-- or should we? Psychological Bulletin, 1970, 74, 68-80.

Gibbons, M. Individualized instruction: A descriptive analysis. New York: Teachers College Press, 1971.

Holland, J. G. Variables in adaptive decisions in individualized instruction. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1975.

Holland, J. G., Solomon, C., Doran, J., & Frezza, D. A. The analysis of behavior in planning instruction. Reading, Mass.: Addison-Wesley, in press.

Holt, J. How children fail. New York: Dell Publishing Company, 1964.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. Inequality: A reassessment of the effect of family and schooling in America. New York: Basic Books, 1972.

Kenny, D. A. A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. Psychological Bulletin, 1975, 82, 345-362.

Leinhardt, G. Evaluation of the educational process in individualized instruction: Third year report. Paper presented at the meeting of the American Educational Research Association, Washington, D. C., March-April 1975.

Lohnes, P. R. Statistical descriptors of school classes. American Educational Research Journal, 1972, 9, 547-556.

Overall, J. E. A survey of issues related to the analysis of observational data in longitudinal research. Galveston, Texas: University of Texas Medical Branch, 1975.

Richards, J. M., Jr. A simulation study of the use of change measures to compare educational programs. American Educational Research Journal, 1975, 12, 299-311.

Tukey, J. W. Exploratory data analysis (3 vols.). Philippines: Addison-Wesley, 1970. (Limited Preliminary Edition)

Tukey, J. W. Some graphic and semigraphic displays. In T. A. Bancroft & S. A. Brown (Eds.), Statistical papers in honor of George W. Snedecor. Ames, Ia.: Iowa State University Press, 1972.

Veldman, D. J. COMMAP: A computer program for commonality analysis. Unpublished manuscript, University of Texas at Austin, 1975.

White, S. H., Day, M. C., Freeman, P. K., Hantman, S. A., & Messenger, K. P. Federal programs for young children: Review and recommendations (Vol. 4). Washington, D. C.: U. S. Government Printing Office, 1973.

Wiley, D. E. Another hour, another day: Quantity of schooling, a potent path for policy. In H. W. Sewell, R. M. Hauser, & D. L. Featherman (Eds.), Schooling and achievement in American society. New York: Seminar, 1973.

APPENDIX A

IN-CLASS INTERVIEWING

Page

In-Class Teacher Interview Questionnaire A-1

In-Class Teacher Interview Questionnaire: Directions for Obtaining Measures A-14

In-Class Teacher Interview Questionnaire: Directions for Coding. A-27

In-Class Teacher Interview Questionnaire

Name of Interviewer ____________________

Date ____________________

School District ____________________

School ____________________

Grade ____________________

Teacher's Name ____________________

Years of Teaching Experience ____________________

1. How many students are enrolled in this classroom? ____________________
2. How many students are present today? ____________________
3. What is the average yearly attendance?

   $\frac{\#\text{ student days}}{\#\text{ students}}$ ____________________
4. How many school days are there in the year?
5. How many children have transferred into this room? ____________________
6. How many children have transferred out of this room? ____________________
7. How many adults are usually in this room? ____________________
8. How many minutes are in the school day? (8e - 8b - [8d - 8c]) ____________________

   8a. What time do students arrive? ____________________

   8b. What time do classes begin? ____________________

   8c. What time are students dismissed for lunch? ____________________

   8d. What time do students return from lunch? ____________________

   8e. What time are students dismissed for the day? ____________________

9. How many minutes a day are spent in mathematics instruction? ______________________

10. How many minutes a day are spent in reading or reading-related instruction? ______________________

11. How much homework is assigned in mathematics?

None				A Lot
1	2	3	4	5

12. How much homework is assigned in reading?

None				A Lot
1	2	3	4	5

13. How frequently is peer tutoring used in this classroom?

Never				Very Frequently
1	2	3	4	5

14. Are games and/or contests used in the teaching of math or reading in this classroom?

Never				Very Frequently
1	2	3	4	5

15. Do students score or grade their own tests?

Never				Very Frequently
1	2	3	4	5

16. Do students decide how they want to work--alone or in small groups or teams?

Never				Very Frequently
1	2	3	4	5

17. Do students decide whom to sit next to in math or reading?

Never				Very Frequently
1	2	3	4	5

18. Rate the reading materials generally available in the classroom as to how interesting the students find them.

Not Interesting				Very Interesting
1	2	3	4	5

19. Rate the mathematics materials generally available in the classroom as to how interesting the students find them.

Not Interesting				Very Interesting
1	2	3	4	5

20. Within a single mode of reading instruction such as pencil and paper, rate the variety of instruction format.

No Variation				Large Amount of Variation
1	2	3	4	5

21. Within a single mode of mathematics instruction such as pencil and paper, rate the variety of instruction format.

No Variation				Large Amount of Variation
1	2	3	4	5

22. Check the various modes of instruction in use in reading.

Audio tapes ________________
Workbooks or sheets ________________
Other books ________________
Film strips ________________
Additional (specify) ____________________________
____
____

23. Check the various modes of instruction in use in mathematics.

Workbooks or sheets ________________
Flash cards ________________
Games ________________
Manipulatives ________________
Audio tapes ________________
Film strips ________________
Additional (specify) ____________________________
____
____

24. On the average, how long after a student has done a specific academic task does he/she receive information about the correctness of his/her performance?

		How Many
Minutes	____________	________
Hours	____________	________
Days	____________	________

25. What types of objectives are present in the reading curriculum? (check one)

General _____
Specific _____
Both general and specific _____
No objectives at all _____

26. If any specific objectives are present, rate the reading curriculum on the clarity of its specific objectives.

Unclear				Clear
1	2	3	4	5

27. When a new objective is presented in the reading materials, which of the following appear? (check appropriate lines)

A written statement of the objective _____
An example of the new objective _____
A range of examples of the new objective _____
Practice problems for the student to complete that are different from the example _____
A non-example _____

28. Rate the reading curriculum as to how closely the available materials match the stated objectives.

No Match				Close Match
1	2	3	4	5

29. At the beginning of the year, how do you decide where a child should start in reading or in which reading group he/she belongs? (check appropriate lines)

By starting at the beginning of the text or series _____
By the results of a standardized test _____
By the results of a curriculum-designed test _____
By the results of a test that you developed _____
From the placement of the child at the end of the previous year _____
From the wishes of the child _____
Other, please specify _______________________________________

30. On the average, how frequently do you think a child starts on a new reading objective (topic)? (check one)

Once a day _____
Twice a week _____
Once a week _____
Once every other week _____
Once a month _____

31. On the average, what amount of reading time do you think a child spends reviewing material that he/she has already learned? _______

32. What types of objectives are present in the mathematics curriculum?

General _____
Specific _____
Both general and specific _____
No objectives at all _____

33. If any specific objectives are present, rate the mathematics curriculum on the clarity of its specific objectives.

Unclear				Clear
1	2	3	4	5

34. When a new objective is presented in the mathematics materials, which of the following appear? (check appropriate lines)

A written statement of the objective _____
An example of the new objective _____
A range of examples of the new objective _____
Practice problems for the student to complete that are different from the example _____
A non-example _____

35. Rate the mathematics curriculum as to how closely the available materials match the stated objectives.

No Match				Close Match
1	2	3	4	5

36. At the beginning of the year, how do you decide where a child should start in mathematics? (check appropriate lines)

By starting at the beginning of the text or series _____
By the results of a standardized test _____
By the results of a curriculum-designed test _____
By the results of a test that you developed _____
From the placement of the child at the end of the previous year _____
From the wishes of the child _____
Other, please specify ___________________________________

37. On the average, how frequently do you think a child starts on a new mathematics objective (topic)? (check one)

Once a day _____
Twice a week _____
Once a week _____
Once every other week _____
Once a month _____

38. On the average, what amount of mathematics time do you think a child spends reviewing material that he/she has already learned? __________

39. Is there a systematic way of assessing student initial abilities built into the reading curriculum (e.g., are some types of placement tests provided or is a procedure recommended)? __________

40. Do you use it? ________

41. Is there an informal way that you use to assess student initial abilities in reading? ______________

42. If so, please specify. ______________________________

43. Is there a systematic way of assessing student mastery of specific skills built into the reading curriculum (e.g., are tests provided)? ______________

44. Do you use it? ____________

45. Rate the reading curriculum according to how clear it is what the next unit should be at the completion of a unit of material.

Unclear				Very Clear
1	2	3	4	5

46. How frequently do you skip around in the reading sequence or text?

Never				Often
1	2	3	4	5

47. How frequently do you create your own reading materials in order to improve the sequencing that is specified in the reading curriculum?

Never				Often
1	2	3	4	5

48. Is there a systematic way of assessing student initial abilities built into the mathematics curriculum (e.g., are some types of placement tests provided or is a procedure recommended)?

____

49. Do you use it? ___________

50. Is there an informal way that you use to assess student initial abilities in mathematics? ______________

51. If so, please specify. ______________________________

52. Is there a systematic way of assessing student mastery of specific skills built into the mathematics curriculum (e.g., are tests provided)? ____________________

53. Do you use it? ______________

54. Rate the mathematics curriculum according to how clear it is what the next unit should be at the completion of a unit of material.

Unclear				Very Clear
1	2	3	4	5

55. How frequently do you skip around in the mathematics sequence or text?

56. How frequently do you create your own mathematics materials in order to improve the sequencing that is specified in the mathematics curriculum?

Never				Often
1	2	3	4	5

57. Who decides what skill or concept the student will work on in reading? (check most appropriate lines)

Parents _____
Teacher _____
Curriculum _____
Child _____

58. Who decides what skill or concept the student will work on in mathematics? (check most appropriate lines)

Parents _____
Teacher _____
Curriculum _____
Child _____

59. If a student does not pass a test in reading, what do you usually do? (check appropriate lines)

Tutor _____
Give special work _____
Continue on _____
Give extra homework _____
Repeat the material covered _____
Other, please specify ____

60. If a student does not pass a test in mathematics, what do you usually do? (check appropriate lines)

Tutor _____
Give special work _____
Continue on _____
Give extra homework _____
Repeat the material covered _____
Other, please specify ____

61a. If a student has been working for several days on a difficult concept or skill in reading and seems to be making no progress toward mastery, what do you do? ____________________________

____

61b. If a student has been working for several days on a difficult concept or skill in mathematics and seems to be making no progress toward mastery, what do you do? ____________________________

____

62. Are the children self-paced in reading?

Yes _____
No _____

63. Are the children self-paced in mathematics?

Yes _____
No _____

64. Over the last two months, what is the largest and smallest number of units completed in reading?

Largest	Smallest
_______	_______

65. Over the last two months, what is the largest and smallest number of units completed in mathematics?

Largest	Smallest
_______	_______

66. In reading, in what size group are children receiving instruction?

Whole class _____
Individual _____
Small group _____
How many groups are there? _____
How many children are in each group? _____
How often are groups re-formed? _____

67. In mathematics, in what size group are children receiving instruction?

Whole class _____
Individual _____
Small group _____
How many groups are there? _____
How many children are in each group? _____
How often are groups re-formed? _____

68. List the assignments by student in reading for today.

Today's Date ____________

I.D.	Book or Text Series	Today's Assignment Pages	Appropriate Identifying Information			Last Test Given	Is This a Diagnostic Test?	How Many Items Did the Student Get Correct out of the Total	Is This a "Passing" Score?
			Level	Unit	Skill				
	Workbook-McGraw-Hill	12-20	B	2	5	9/24/76	Yes	6/10	Yes

For use by coder of questionnaire only.

Reading

a. Percent of unique assignments ________

b. Average number of days since the last test ________

c. Range of days since the last test ________

d. Average percent correct on the last test ________

69. List the assignments by student in mathematics for today.

Today's Date ________________

I.D.	Book or Text Series	Today's Assignment Pages	Appropriate Identifying Information			Last Test Given	Is This a Diagnostic Test?	How Many Items Did the Student Get Correct out of the Total	Is This a "Passing" Score?
			Level	Unit	Skill				
	Addison-Wesley	8, 9	A	9	4	9/27/76	Yes	6/10	Yes

For use by coder of questionnaire only.

Mathematics

a. Percent of unique assignments ________

b. Average number of days since the last test ________

c. Range of days since the last test ________

d. Average percent correct on the last test ________

Overlap Data. The following information will be used in calculating the curriculum overlap. It should be collected in the fall.

What is the name(s) of the math series being used? ____________________

Are other texts being used? ____________________

If more than one series is being used, report names of all published series. ____________________

Do special groups use them (e.g., advanced or remedial groups)? ____________________

If so, who uses them? ____________________

Do you use the extra series to "patch" gaps in the existing one? ____________________

List the topics that are taught using supplementary materials (e.g., division of fractions, the number line). ____________________

What is the name(s) of the reading series being used? ____________________

Are other texts being used? ____________________

If more than one series is being used, report names of all published series. ____________________

Do special groups use them (e.g., advanced or remedial groups)? ____________________

If so, who uses them? ____________________

Do you use the extra series to "patch" gaps in the existing one? ____________________

List the topics that are taught using supplementary materials. ____________________

The following information will also be used in calculating the curriculum overlap. It should be collected in the spring.

70. Teacher's estimated percentage of overlap in reading ___________

71. Teacher's estimated percentage of overlap in mathematics ______

In-Class Teacher Interview Questionnaire:

Directions for Obtaining Measures

1. Obtain this information from the number of students on the teacher's permanent and updated attendance roster.

   NOTE: Count the names; do not use self-reports.

2. Count the number of students present in the room, and count the number on the attendance sheet. If the two numbers differ, try to get clarification on the difference (e.g., Johnny went to tutoring) and record the maximum.

3. From the end-of-year attendance summaries, obtain the total number of school days attended by each child and add them. Divide this sum by the total number of students.

   Include: All children who were present in the fall and spring.

   Exclude: Any child who was absent in either the fall or spring and was also absent for more than 40% of the school year.

   Example: Child A starts school on November 10 and is present every day thereafter--include. Child B starts school on September 1, is enrolled through the end of the school year, and attends less than 50 days--include. Child C starts school on September and transfers 25 days later--exclude.

   NOTE: Do not ask this question in the fall.

4. Obtain this information from a school administrator or teacher. (Include only days that children are in school--exclude in-service days, etc.)

5-6. Obtain this information from end-of-year records of the teacher or principal.

   NOTE: Do not ask this question in the fall.

7. Ask the teacher, count the adults, and discuss any differences. Include student teachers if they are present more than 50% of the year (e.g., all year, 3 days a week).

8. Ask the teacher Items 8a-8e. Observe 8a-8c and resolve any differences. Calculate 8 from the set of information obtained.

9-10. Ask these questions directly to the teacher.

11-12. Ask these questions directly to the teacher.

None (1) means that no homework is assigned. A lot (5) means that homework is assigned every night and at least equals the amount of time spent in the subject during school hours.

13. Ask this question directly to the teacher.

Definition: Peer tutoring is occurring if one student in the class teaches another student(s) in the same class a specific activity or helps him/her to rehearse a specific skill. Look around the room and see if there are any groups of two or more students working together. Ask the teacher if peer tutoring ever occurs. If you saw no evidence of it and the teacher says "no," record 1. If you saw no evidence and the teacher says "yes," ask when it occurs (e.g., subject) and ask for the most recent (or immediate future) specific case of one student tutoring another. If the answers are vague and completely nonspecific, record 1, but if the answers are specific and reasonable, ask the teacher to rate how frequently peer tutoring occurs. If you saw evidence of possible peer tutoring and the teacher says "yes," recall the incidents you saw. Ask if they were, in fact, cases of peer tutoring. If so, ask for a rating; if not, probe as above. If you saw evidence of possible peer tutoring and the teacher says "no," confirm this response by describing what you saw and asking what it was.

14. Ask this question directly to the teacher.

In this question and the next three questions, 1 means never and 5 means every day for both reading and math. Look around the room. If you see any use of games or contests, ask for a rating. If you don't, ask the teacher. If he/she says "no," record 1. If he/she says "yes," ask for specific examples and when they were used, then ask for a rating. If he/she is vague or unclear, mark 1.

15-17. Ask these questions directly to the teacher. Use the same technique as in Item 14.

18-19. After reading these questions once, explain the following:

In asking you to respond to these questions, we are trying to elicit your judgment about how motivating the curricula are to the students you teach. While we can judge this from a theoretical point of view, you have invaluable practical information about how much the children are interested or intrigued by the material contained in the text. Please rate the curriculum using your own experience and considering such things as pacing, format, artistry, etc. If, in addition, you would like to comment on any part of the curriculum, please feel free to do so. Your comments will be read and valued.

20-21. After reading these questions once, explain the following:

For these questions, we are asking you to make a similar judgment. However, for the ratings, consider the main mode(s) of presentation such as textbook or workbook, and consider the variation that occurs within that mode. Consider variation in patterns of presentation, layout, and type of response required by the student. For example, a program that presents text followed by questions for the majority of the instruction has little variation and would be considered a 1. On the other hand, a program that has text followed by a variety of activities, such as writing short answers, writing plays, and writing questions for other students to answer, has large variation in mode and would be considered a 5.

22-23. Check all modes of instruction that are evident in the classroom and then ask the teacher for specifics about the use of additional modes.

24. Ask this question directly to the teacher.

Observe actual correction practices and resolve any differences. See, for example, if students' work is corrected while they are doing it or immediately afterward. If so, give your answer in minutes and disregard the other time units. If classwork is returned during the period, check to see when it was turned in to the teacher. If it was turned in the previous day, record the number of days "old" the work is.

25. After reading this question once, explain the following:

A general objective is a global, program goal.

Examples of general reading objectives:

1. The child will become an independent reader, choosing to read for pleasure.
2. The child will be able to extract meaning from written text.

Use only one example on the first reading. The second example should be read only if the information is not clear to the teacher.

A specific objective is one that is linked with the instructional materials and is most often in behavioral terms.

Examples of specific reading objectives:

1. The child will be able to decode all one-syllable words in his/her reading text.
2. The child will be able to read a short paragraph story (three to four sentences) and paraphrase the story.

Again, use only one example on the first reading. The second example should be read only if the information is not clear to the teacher.

Then ask the teacher if the reading curriculum has general objectives only, specific objectives only, both general and specific objectives, or no objectives at all.

26. If no specific objectives are present in the reading curriculum, skip to Item 27. If there are some specific objectives in the reading curriculum, ask the following question:

Consider the clarity of the specific objectives in the reading curriculum. On a scale from 1 to 5, rate the clarity of the specific objectives.

A clear objective is one that is focused and narrow. It precisely states what the child will know how to do on completion of the materials. An example of a clear specific objective that would be considered a 5 is: upon completing this unit, the student can pronounce all one-syllable words of the CVC form.

An unclear objective is one that is dispersed and broad. It does not describe precisely what the child will be able to do on completion of the materials. An example of an unclear specific objective that would be considered a 1 is: upon completing this unit, the child knows how to read one-syllable words.

27. After reading this question once, explain the following:

A written statement of a reading objective may appear either in the teacher's manual or the child's text or in both. An example of a written statement of a reading objective is: the child will be able to pronounce any one-syllable word with two vowels, the final vowel being a silent "e."

An example of the new objective is: to present the child with a list of words that conform to this rule (sale, bale, male) and to pronounce the words for him/her.

A range of examples covers the broadest extremes of the domain. An example is: sale, Joe, ate.

Practice problems for the objective could be a list of words conforming to the rule that the child must pronounce by himself/herself.

A non-example for the objective is one outside of the domain. An example is: rut, not, near.

28. Ask this question directly to the teacher.

The materials match closely (5) if the contents of both the materials and the objectives are identical. The materials do not match the objectives (1) if the contents are not identical.

29. Ask this question directly to the teacher. If he/she gives multiple answers or the answer is vague, approach it in reverse:

Ask "Who is in the top reading group?" Pick one of the list of names. Ask "How did Ann get placed in that group?" Repeat the question for two or three other children until you have the most commonly used criteria for grouping.

NOTE: Do not ask this question in the spring.

30. Ask this question directly to the teacher.

If the teacher is unclear as to the meaning of an objective, define it: An objective is a single block of material or a topic that a child can learn in a fairly short period of time. It is often, but not always, followed by a test.

If the teacher is still unclear, change the question so that a specific child is mentioned: "What is Althea working on? When did she start that? When will she finish?"

31. Ask this question directly to the teacher. If the teacher is uncertain as to what is meant, explain the following:

A student may, for example, work on objectives four out of five days and on the fifth day review skills that he/she has already learned. If this were the case, then one-fifth of the time would be spent on review and the amount of time recorded would be the length of one subject period. Another student may spend the first 10 minutes of a 30-minute period reviewing previously learned objectives. In this case, the amount of time recorded would be 10 minutes.

Then ask the teacher to try again to estimate the amount of scheduled time that students, on the average, spend doing some form of review such as practice or drill.

32. After reading this question once, explain the following:

A general objective is a global, program goal.

Examples of general mathematics objectives:

1. The child will demonstrate problem-solving skills.
2. The child will enjoy playing mathematical games.

Use only one example on the first reading. The second example should be read only if the information is not clear to the teacher.

A specific objective is one that is linked with the instructional materials and is most often in behavioral terms.

Examples of specific mathematics objectives:

1. The child will be able to add two one-digit numbers.
2. The child will be able to count to 100 by fives.

Again, use only one example on the first reading. The second example should be read only if the information is not clear to the teacher.

Then ask the teacher if the mathematics curriculum has general objectives only, specific objectives only, both general and specific objectives, or no objectives at all.

33. If no specific objectives are present in the mathematics curriculum, skip to Item 34. If there are some specific objectives in the mathematics curriculum, ask the following question:

As you did for reading, consider for mathematics the clarity of the specific objectives. On a scale from 1 to 5, rate the clarity of the specific objectives.

A clear specific objective that would be considered a 5 is: given a series of one-digit, two-numeral addition problems in the vertical form, the child will be able to correctly calculate their sums.

An unclear specific objective (1) might be: given a set of one-digit addition problems, the child will understand how to find their sums.

34. After reading this question once, explain the following:

A written statement of a mathematics objective may appear either in the teacher's manual or the child's text or in both. An example of a written statement of a mathematics objective is: the student will be able to sum any two one-digit whole numbers.

An example of the new objective is: to present the child with a list of solved problems and conceptual methods for calculating correct answers to the problems (e.g., number line representations of the solutions to 3+1=4, 4+1=5, 1+2=3).

A range of examples covers the broadest extremes of the domain. An example is: 5+4=9, 8+9=17, 6+7=13.

Practice problems for the objective could simply be a list of unsolved problems for the student to complete.

A non-example for the objective is one outside of the domain. An example is: 1-1=?

35. Ask this question directly to the teacher.

The materials match closely (5) if the contents of both the materials and the objectives are identical. The materials do not match the objectives (1) if the contents are not identical.

36. Ask this question directly to the teacher. If he/she gives multiple answers or the answer is vague, approach it in reverse:

Ask "Who is in the top mathematics group?" Pick one of the list of names. Ask "How did Ann get placed in that group?" Repeat the question for two or three other children until you have the most commonly used criteria for grouping.

NOTE: Do not ask this question in the spring.

37. Ask this question directly to the teacher.

If the teacher is unclear as to the meaning of an objective, define it: An objective is a single block of material or a topic that a child can learn in a fairly short period of time. It is often, but not always, followed by a test.

If the teacher is still unclear, change the question so that a specific child is mentioned: "What is Althea working on? When did she start that? When will she finish?"

38. Ask this question directly to the teacher. If the teacher is uncertain as to what is meant, explain the following:

A student, for example, may work on new objectives four out of five days and on the fifth day review skills that he/she has already learned. If this were the case, then one-fifth of the time would be spent on review and the amount of time recorded would be the length of one subject period. Another student may spend the first 10 minutes of a 30-minute period reviewing previously learned objectives. In this case, the amount of time recorded would be 10 minutes.

Then ask the teacher to try again to estimate the amount of scheduled time that students, on the average, spend doing some form of review such as practice or drill.

39-40. Ask Item 39 directly to the teacher.

Record "yes" or "no." If "yes," then ask if this is the way most children are placed in the curriculum. If so, record "yes" for Item 40 and retain "yes" for Item 39. If the answer for Item 40 is "no," record "no" for Item 40 and retain "no" for Item 39 or change the "yes" in Item 39 to "no" by lightly crossing out the "yes" and writing "no."

41-42. Ask Item 41 directly to the teacher.

If the teacher is uncertain, say: "Sometimes teachers don't always rely on the regular placement tests or they don't have any placement tests to begin with. Do you use some other way of deciding where to start the child?" Then ask for the "specific method" in order to verify the answer to Item 41. If there are no specific methods, record the answer to 41 as "no." If there are specific methods, record the answer as "yes."

43-44. Ask Item 43 directly to the teacher.

Record "yes" or "no." If "yes," then ask if this is the way in which it is decided that most children have mastered a particular skill. If so, record "yes" for Item 44 and retain "yes" for Item 43. If the answer for Item 44 is "no," record "no" for Item 44 and retain "no" for Item 43 or change the "yes" in Item 43 to "no" by lightly crossing out the "yes" and writing "no."

45. Ask this question directly to the teacher.

It is very clear what the next unit should be (i.e., rated 5) if the curriculum provides specific sequencing information. For example, the curriculum might list the following sequence: If a student(s) passes Unit 1, level C, he/she should go to Unit 1, level D. If a student(s) does not successfully master Unit 1, level C, he/she should go to Unit 1, level C alternate.

Although the curriculum may provide little or no sequencing information, the way in which the units of material are lettered or numbered in itself may provide clear information. For example, it may be obvious that after Unit 1 is completed, Unit 2 is the unit that should be covered next. The rating in this case would also be 5.

It is unclear what the next unit should be (i.e., rated 1) if the curriculum does not provide adequate sequencing information. For example, at the end of Unit 1, level C, the curriculum might state: If a student(s) has not mastered the skills in this unit, review those skills before continuing on.

It is also not clear what the next unit should be if units may be covered in any sequence and no guidance is provided.

46. Ask this question directly to the teacher.

This question would be rated 5 (often) if the teacher rarely follows the sequence provided and possibly skips some material totally. The rating 1 (never) would be used if the teacher always follows the prescribed routing pattern or takes each unit in the order in which it is presented.

47. Ask this question directly to the teacher.

This question would be rated 5 (often) if the teacher constructs his/her own materials every day to improve sequencing and 1 if he/she never does.

48-49. Ask Item 48 directly to the teacher.

Record "yes" or "no." If "yes," then ask if this is the way most children are placed in the curriculum. If so, record "yes" for Item 49 and retain "yes" for Item 48. If the answer for Item 48 is "no," record "no" for Item 49 and retain "no" for Item 48 or change the "yes" in Item 48 to "no" by lightly crossing out the "yes" and writing "no."

50-51. Ask Item 50 directly to the teacher.

If the teacher is uncertain, say, "Sometimes teachers don't always rely on the regular placement tests or they don't have any placement tests to begin with. Do you use some other way of deciding where to start the child?" Then ask for the "specific method" in order to verify the answer to Item 50. If there are no specific methods, record the answer to 50 as "no." If there are specific methods, record the answer as "yes."

52-53. Ask Item 52 directly to the teacher.

Record "yes" or "no." If "yes," then ask if this is the way in which it is decided that most children have mastered a particular skill. If so, record "yes" for Item 53 and retain "yes" for Item 52. If the answer for Item 53 is "no," record "no" for Item 53 and retain "no" for Item 52 or change the "yes" in Item 52 to "no" by lightly crossing out the "yes" and writing "no."

54. Ask this question directly to the teacher.

It is very clear what the next unit should be (i.e., rated 5) if the curriculum provides specific sequencing information. For example, the curriculum might list the following sequence: If a student(s) passes Unit 1, level C, he/she should go to Unit 1, level D. If a student(s) does not successfully master Unit 1, level C, he/she should go to Unit 1, level C alternate.

Although the curriculum may provide little or no sequencing information, the way in which the units of material are lettered or numbered in itself may provide clear information. For example, it may be obvious that after Unit 1 is completed, Unit 2 is the unit that should be covered next. The rating in this case would also be 5.

It is unclear what the next unit should be (i.e., rated 1) if the curriculum does not provide adequate sequencing information. For example, at the end of Unit 1, level C, the curriculum might state: If a student(s) has not mastered the skills in this unit, review those skills before continuing on.

It is also not clear what the next unit should be if units may be covered in any sequence and no guidance is provided.

55. Ask this question directly to the teacher.

This question would be rated 5 (often) if the teacher rarely follows the sequence provided and possibly skips some units of material totally. The rating 1 (never) would be used if the teacher always follows the prescribed routing pattern or takes each unit in the order in which it is presented.

56. Ask this question directly to the teacher.

This question would be rated 5 (often) if the teacher constructs his/her own materials every day to improve sequencing and 1 if he/she never does.

57-58. Ask these questions directly to the teacher.

Try to get at the most prevalent form of decision making. Try not to check all four lines.

59-60. Ask these questions directly to the teacher.

61a-61b. Ask these questions directly to the teacher.

62-63. After reading these questions once, explain the following:

Self-paced means that the child's mastery of and/or interest in specific material determines when he/she goes on to new material.

If a child is not self-paced, he/she is usually in a classroom where progression to new material is decided by the teacher when the majority of the class has completed a unit of material.

64-65. Item 64 is to be asked only if the teacher answered "yes" for Item 62.

Item 65 is to be asked only if the teacher answered "yes" for Item 63.

For both questions, ask the teacher which child has completed the greatest number of units in the reading (or mathematics) text over the last two months. The teacher will most likely need to consult his/her records. Then ask the teacher which child has completed the fewest number of units in reading (mathematics) over the last two months. Again, the teacher will most likely consult the records.

Record only the number of units, not the child's name.

66-67. Ask these questions directly to the teacher. Try to obtain only one answer for the type of grouping. If the answer is "small group," ask for information on number, size, and re-forming of groups.

68-69. The purpose of these items is to estimate the uniqueness and distribution of students' assignments and testing. Uniqueness means how many different assignments are given in a class on any one day. If each student receives a different assignment, then that classroom would have 100% unique assignments. On the other hand, if all students in a class of 25 receive the same assignment, then only one assignment is unique and the percentage is 4%. How often students are tested and how well they do on a given test give an estimate of the monitoring procedures and matching of the student and the curriculum.

The date that you go into the classrom to record the information must be noted in the upper right-hand corner next to "Today's date ______." This date will be used later in coding the responses on the questionnaire. It must match the actual day you collect the information for each classroom.

In order to ensure that all student assignments are listed, list each student in the "I. D." column. Using a class roster or attendance sheet, list the initials of each child who is present that day. If two children have the same initials, include more information to distinguish between them, such as a middle initial. If a student is absent on that day but has been in the classroom on the previous day, include him/her in the I. D. list. Students who are not listed should include only those who have been absent for two or more days, moved, or transferred to another class.

In order to determine the assignment for each student, ask the teacher how students are assigned classwork or homework for that day in reading and mathematics. Since Items 68 and 69 require identical information, only Item 68 (reading assignments) will be discussed here. Use the same procedures, however, to complete Item 69 (mathematics assignments).

There are several ways in which a student's assignments may be recorded. If each student has an individual assignment sheet or prescription card, match it with the student's initials and list the day's assignment as specifically as possible. Note the name of the book or text series in which the student is working, the page numbers assigned, and identifying information about the level or chapter, unit, and skill or concept. If the teacher keeps a log of each child's assignments, the same procedure may be used by matching each child and listing the assignments. If all children receive the same assignment for the day, list all the identifying information for that assignment and write the number of children counted under the I. D. column. If students are assigned work by groups, list the assignment and the number of children in each group. If all or some of the students are sent to a different class for reading, go to that room and follow the same procedure for determining their assignment. If any of the assignments are unclear or you cannot find some of the information, use the teacher as a resource for clarification. Try to work quickly and cause as little disruption as possible. When you are finished, each student should be accounted for with specific assignments listed.

If the text is only broken down into chapters, list other pertinent information such as "two-digit addition." An example of this might be:

I. D.	Book	Pages	Level	Unit	Skill
AMS	Scott-Foresman	26-31	2	Plurals	"ing" endings

The final four columns are used to record for each child the date of the last test given, if that test is a diagnostic one, the number of items that child had correct out of the total number of items, and if the score is passing. At the beginning of the school year, you should request that teachers keep logs of this information. The questions on the bottom of pages A-10 and A-11 will be calculated by the coder from the information that you provide on the forms.

Overlap Data. Ask the principal at the time of initial contact. Ask the teacher at the time of interviewing. Include supplementary texts if the teacher reports using any. If the principal and teacher report different texts, try to decide which is really being used (check student desks, ask students, etc.).

70-71. NOTE: These items are intended to determine those skills taught by the teacher that are also covered by the CTBS battery. These questions should be asked in a final interview with the teacher. The interview should take place immediately after the spring testing is completed or while the testing is in progress. The interviewer can be either the same individual who conducted the previous interviewing or the person administering the spring battery of achievement tests.

Hand the teacher the appropriate form and grade level of the test battery and say:

Please circle all item numbers that you think the majority of your students have been taught either by the curriculum or by you. Look at the format of the items as well as the skills involved before deciding if an item has been taught. We are trying to determine if the test is actually testing what your students have learned. For example, if your students have learned to add 5+3, but this problem has always been presented in vertical form, you would not circle that item when it appears in horizontal form. As another example, if your students have been asked questions after reading a story, but the questions were always taken literally from the text, then a question that requires making an inference from the text would not be circled. Please go through the entire test and read the questions thoroughly before making a decision.

When the teacher has gone through the entire achievement test and circled all appropriate items, put his/her name on the booklet with other identifying information such as date, school, and grade level.

In-Class Teacher Interview Questionnaire:

Directions for Coding

1. Record one number only, which is the average of the number of students enrolled in the fall and the number of students enrolled in the spring. If this average number is not a whole number, round off to the greater number of children.

2. Calculate a percentage of each questionnaire (fall and spring)-- the number in Item 2 divided by the number in Item 1. Record the average of the percentages over the two passes.
NOTE: "Over the two passes" refers to the fall and spring data collection efforts.

3. Record the average number of days a student attends.
NOTE: On spring questionnaire only.

4. Record a single number. This number should remain constant from fall to spring.

5-6. Record the total numbers.
NOTE: On spring questionnaire only.

7-10. Record the average over the two passes.

11-21. Record the average ratings over the two passes.

22. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Thus, the measure consists of 5 binary codes; e.g., 01000 = only workbooks or worksheets are used in the instruction of reading.
Combine the results over the two passes.

23. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Thus, the measure consists of 7 binary codes; e.g., 0111000 = flash cards, games, and manipulatives are used in the instruction of mathematics. Combine the results over the two passes.

24. Transform all answers to hours; consider each day to be six hours. Thus:

30 minutes = .50
2 days = 12.00
3 hours = 3.00

Record the average over the two passes.

25. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Thus, the measure consists of 4 binary codes; e.g., 0100 = only specific objectives exist. Combine the results over the two passes.

26. Record the average rating over the two passes; record zero (0) if no specific objectives exist.

27. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Combine the results over the two passes.

28. Record the average rating over the two passes. Record zero (0) if there are no objectives.

29. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
NOTE: On fall questionnaire only.

30. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Combine the results over the two passes.

An example of combined results: In the fall, a new reading objective is introduced once a week (code = 00100) and in the spring, a new reading objective is introduced once a month (code = 00001). The combined results are coded 00101.

31. Record the average amount of time per week over the two passes.

32. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Thus, this measure consists of 4 binary codes; e.g., 0100 = only specific objectives exist. Combine the results over the two passes.

33. Record the average rating over the two passes.

34. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Combine the results over the two passes.

35. Record the average rating over the two passes.

36. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
NOTE: On fall questionnaire only.

37. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Combine the results over the two passes.

38. Record average amount of time per week over the two passes.

39. no = 0
yes = 1

Record one (1) if "yes" on either fall or spring pass. Record zero (0) if "no" on both passes.

40. No code--not recorded.

41. no = 0
yes = 1

Record one (1) if "yes" on either fall or spring pass. Record zero (0) if "no" on both passes.

42. No code--not recorded.

43. no = 0
yes = 1

Record one (1) if "yes" on either fall or spring pass. Record zero (0) if "no" on both passes.

44. No code--not recorded.

45-47. Record the average ratings over the two passes.

48. no = 0
yes = 1

Record one (1) if "yes" on either fall or spring pass. Record zero (0) if "no" on both passes.

49. No code--not recorded.

50. no = 0
yes = 1

Record one (1) if "yes" on either fall or spring pass. Record zero (0) if "no" on both passes.

51. No code--not recorded.

52. no = 0
yes = 1

Record one (1) if "yes" on either fall or spring pass. Record zero (0) if "no" on both passes.

53. No code--not recorded.

54-56. Record the average ratings over the two passes.

57-60. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Combine the results over the two passes.

61a-61b. There are several parts to analyzing the answers to these questions. After the fall interview, one coder should make a master list of all unique answers. This list is then used to create another list of no more than five major categories of responses. Code each questionnaire using 1 (yes) or 0 (no) for each category.

Example: Suppose the five major categories generated from all the questionnaires are: tutor, special work, repeat material, continue on, give homework. If a particular teacher says he/she only tutors students who are not progressing toward mastery, you would code his/her answer as 10000.

The major categories generated in the fall should be used again in the spring. So, if the same teacher reports in the spring that he/she gives homework, his/her answer would be coded 00001.

If the results are the same over the two passes, use them; if different, record the combined results (e.g., the combined results of the above answers would be 10001).

62-63. no = 0
yes = 1

Record one (1) if "yes" on either fall or spring pass. Record zero (0) if "no" on both passes.

64-65. Take the largest number and subtract the smallest number to calculate the answer. Average the answers over the two passes.

66-67. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Combine the results over the two passes.

Record the average over the two passes for the number of groups and group size.

Transform the answer to the regrouping question into frequency by month (e.g., once a week=4, once every 2 months=.5).

68-69. There are four responses that must be coded for each subject area, i.e., reading and mathematics. They are: percent of unique assignments, average number of days since the last test, range of days since the last test, and average percent correct on the last test. The procedures for coding reading and mathematics are the same; only reading will be used for explanatory purposes.

The first item (68a) is determined from the assignment identification information: book or text series, assignment pages, level, unit, and skill. Look at the first assignment. Put a 1 next to it. Look at the rest of the assignments and check off (✓) any other assignment that is the same as number 1. Do not check off number 1. Look at the next assignment on the list that is not checked off. Put a 2 next to it. Look at the rest of the assignments and check off any that are the same as number 2. Continue in the same manner to number 3 and so on through all the assignments until each assignment is either numbered or checked. When you are finished, be sure that all of the numbered assignments are unique, i.e., none of them is the same as any other numbered assignment.

Example:

	Book	Pages	Level	Unit	Skill
1.	Scott-Foresman	26-31	B	2	9
2.	Scott-Foresman	40-43	C	1	6
3.	S-F Workbook	10-11	B	3	1
✓	Scott-Foresman	40-43	C	1	6

In this case, three assignments are unique. Using the last numbered assignment, divide by the total number of assignments listed to get the percent of unique assignments. In the above example, 3 out of 4 assignments are unique; therefore, 75% would be the response to Item 68a.

The column "Last Test Given" will supply the response to Items 68b and 68c. For each date, count the number of school days (excluding weekends, holidays, in-service days, etc.) from that date to the date noted in the upper right-hand corner (Today's Date), counting Today's Date as 0. Put the number of days since the last test next to the date of the last test.

Example:

Today's Date 10/4/76

Last Test Given	
9/24/76	(5)
10/1/76	(1)
9/13/76	(15)
10/4/76	(0)

To determine the average number of days since the last test (68b), sum the number of days and divide by the total number of assignments listed. In the example shown previously, the sum of the days is 21; therefore, the average number of days since the last test would be 21/4 = 5.25. To determine the range of days (68c), find the greatest number of days since the last test and subtract the least number of days since the last test. In the above example, the range would be 15-0 or 15.

The column labeled "How Many Items Did the Student Get Correct out of the Total" is used to determine the average percent correct on the last test (68d). For each score, you must determine the percent correct. In other words, if a student scores 6 out of 10 possible points, he/she has scored 60% correct. When you have found the percent correct for each student to two decimal places, sum the percentages and divide by the total number of students.

Example:

How Many Items Did the Student Get Correct out of the Total	Percent Correct
6/10	60.00
8/15	53.33
9/9	100.00
21/25	84.00
7/16	43.75
	Σ=341.08

If the sum of the percentages is 341.08, divide by 5 (the number of assignments listed). Then the average percent correct on the last test would be 68.22%.

The information on diagnostic testing will be used in a behavioral analysis of the curriculum (see Appendix B-2).

Repeat all of the above steps for Item 69 using the assignments listed in mathematics.

Overlap Data. Questions about the text or series being used will not be coded. This information will be used by the curriculum expert to answer Items 1 and 2 on the Curriculum Analysis Questionnaire.

70-71. Count the number of items circled in all reading tests. Reading tests for grade 1 include:

Letter Sounds
Word Recognition I and II
Reading Comprehension
Language I

Reading tests for grade 3 include:

Reading Vocabulary
Reading Comprehension
Language Expression

The sum of all circled items in reading is then divided by the total number of items to determine the percentage of overlap. The total number of reading items for grade 1 is 100. Therefore, if the sum of all circled items is 73, then the teacher's estimated percentage of reading overlap is 73%.

The total number of reading items for grade 3 is 115. If the sum of all circled items is 81, then the overlap estimate is 70.43%.

Count the number of items circled in all math tests. Math tests for grades 1 and 3 include:

Mathematics Concepts and Applications
Mathematics Computation

The achievement battery for grade 1 includes a total of 56 items in mathematics. The battery for grade 3 contains 98 mathematics items.

Sum the number of circled items and divide by the total number of items to determine the percentage of overlap in mathematics. For example, if a first-grade teacher circled 40 items on the math tests, divide 40 by 56, which yields a percentage of 71.43%.

APPENDIX B-1

ANALYZING CURRICULA

Page

Curriculum Analysis Questionnaire B-1

Curriculum Analysis Questionnaire:
Directions for Obtaining Measures B-6

Curriculum Analysis Questionnaire:
Directions for Coding B-17

Analyses of the Reading and Mathematics Tests
of the CTBS (Expanded Edition), Grade 1,
Level B, Form S B-19

Analyses of the Reading and Mathematics Tests
of the CTBS (Expanded Edition), Grade 3,
Level 1, Form T, Book 1 . . . . B-31

Note: Appendix B-2 begins on page B-53.

Curriculum Analysis Questionnaire

Name of Curriculum Expert ______________________

Date ______________________

School District ______________________

School ______________________

Grade ______________________

Teacher's Name ______________________

1. What is the average percentage of overlap in mathematics? ______________________

2. What is the average percentage of overlap in reading? ______________________

3. Rate the reading materials generally available in the classroom as to how interesting the students find them.

Not Interesting				Very Interesting
1	2	3	4	5

4. Rate the mathematics materials generally available in the classroom as to how interesting the students find them.

Not Interesting				Very Interesting
1	2	3	4	5

5. Within a single mode of reading instruction such as pencil and paper, rate the variety of instruction format.

No Variation				Large Amount of Variation
1	2	3	4	5

6. Within a single mode of mathematics instruction such as pencil and paper, rate the variety of instruction format.

No Variation				Large Amount of Variation
1	2	3	4	5

7. What types of objectives are present in the reading curriculum? (check one)

General _____
Specific _____
Both general and specific _____
No objectives at all _____

8. If any specific objectives are present, rate the reading curriculum on the clarity of its specific objectives.

Unclear				Clear
1	2	3	4	5

9. When a new objective is presented in the reading materials, which of the following appear? (check appropriate lines)

A written statement of the objective _____
An example of the new objective _____
A range of examples of the new objective _____
Practice problems for the student to complete that are different from the example _____
A non-example _____

10. Rate the reading curriculum as to how closely the available materials match the stated objectives.

No Match				Close Match
1	2	3	4	5

11. Which of the following assessment procedures does the reading curriculum include?

	Specific Method
Placement procedure _____	_____
Monitoring procedure _____	_____
Method for assessing mastery of skills _____	_____

12. Rate the reading curriculum as to the degree to which the student and the curriculum can be easily matched.

Difficult to Match				Easy to Match
1	2	3	4	5

13. For reading, rate the degree tc which items on the placement test reflect the curriculum content.

No Relation				High Relation
1	2	3	4	5

14. For reading, rate the degree to which items on the mastery test reflect the curriculum content.

No Relation				High Relation
1	2	3	4	5

15. What types of objectives are present in the mathematics curriculum? (check one)

General _____
Specific _____
Both general and specific _____
No objectives at all _____

16. If any specific objectives are present, rate the mathematics curriculum on the clarity of its specific objectives.

Unclear				Clear
1	2	3	4	5

17. When a new objective is presented in the mathematics materials, which of the following appear? (check appropriate lines)

A written statement of the objective _____
An example of the new objective _____
A range of examples of the new objective _____
Practice problems for the student to complete
    that are different from the example _____
A non-example _____

18. Rate the mathematics curriculum as to how closely the available materials match the stated objectives.

No Match				Close Match
1	2	3	4	5

19. Which of the following assessment procedures does the mathematics curriculum include:

	Specific Method
Placement procedure _____	_______________
Monitoring procedure _____	_______________
Method for assessing mastery of skills _____	_______________

20. Rate the mathematics curriculum as to the degree to which the student and the curriculum can be easily matched.

Difficult to Match				Easy to Match
1	2	3	4	5

21. For mathematics, rate the degree to which items on the placement test reflect the curriculum content.

No Relation				High Relation
1	2	3	4	5

22. For mathematics, rate the degree to which items on the mastery test reflect the curriculum content.

No Relation				High Relation
1	2	3	4	5

For the following set of questions, please respond to each question for reading and for mathematics.

		Reading	Mathematics
23a.	Is there a systematic way of assessing student initial abilities built into the curriculum (e.g., are some types of placement tests provided)?	_______	_______
23b.	Please specify the method of assessing student initial abilities.	_______	_______
24a.	Is there a systematic way of assessing student mastery of specific skills built into the curriculum (e.g., are tests provided)?	_______	_______
24b.	Please specify the method of assessing student mastery.	_______	_______

25. Rate the reading curriculum according to how clear it is what the next unit should be at the completion of a unit of material.

Unclear				Very Clear
1	2	3	4	5

26. Rate the mathematics curriculum according to how clear it is what the next unit should be at the completion of a unit of material.

Unclear				Very Clear
1	2	3	4	5

27. If a student does not pass a test, what does the reading curriculum suggest doing?

Tutor _____
Give special work _____
Continue on _____
Give extra homework _____
Repeat the material covered _____
Other, please specify ______________________________

28. If a student does not pass a test, what does the mathematics curriculum suggest doing?

Tutor _____
Give special work _____
Continue on _____
Give extra homework _____
Repeat the material covered _____
Other, please specify ______________________________

29. What is the principal sequencing system in use in the reading curriculum?

Unordered _____
Linear _____
Branched _____
Modular _____

30. What is the principal sequencing system in use in the mathematics curriculum?

Unordered _____
Linear _____
Branched _____
Modular _____

Curriculum Analysis Questionnaire: Directions for Obtaining Measures

1-2. The purpose of these measures is to assess the degree to which a particular curriculum and an achievement test overlap. In order to estimate these measures, match the test to the curriculum by finding the percentage of the items on the test that are covered by the curriculum. When this information is combined with the amount of the curriculum covered during the school year, an estimate can be made of the opportunity that a group of students has to learn the information tested on the criterion.

Before trying to match an achievement test with a curriculum, scan both in order to gain some familiarity with the material. If a teacher's annotated edition of the text is available, look at such things as the table of contents, goals and objectives, and general layout of the work presented. Also look at the publisher's scope and sequence chart if it is available. For the achievement test being used, look at the technical manual item descriptors, the subtests, the kinds of items included, the format of the items, and the directions for administering the test.

A listing of the item types and formats used in the CTBS is presented later in this Appendix. This listing includes information concerning the skill or concept being tapped (e.g., addition facts less than 10) as well as the form of presentation (e.g., word problem, horizontal). Using it as a guide, begin to move through the curriculum. It is probably easiest to start at the beginning of the text, that is, with Chapter 1, Unit 1, or similar subdivisions of the curriculum. As each item is located in the curriculum materials, mark the chapter or unit in which the item is first covered (see Example 1, page B-7). The item may be covered later as review or practice; however, only the initial appearance of that item should be listed. After you complete the first unit of the curriculum, sum the number of items covered in that unit and calculate the percentage of the total number of items in the test. An example for a test containing a total of 100 items is:

Unit	Items Covered	% of Total
Chapter 1	6	6%

Work through each unit in the same fashion by marking the unit or chapter in which a test item is covered and by noting the number and percentage of total test items covered within that unit.

Example 1

Item	Format	No. Numerals and No. Digits	Skill(s) Required	Unit of Material
1	vertical	3, 1 digit	10 < sum < 20, using 0 as an addend	Chapter 2
2	horizontal	[illegible], 1 digit	10 < sum < 20	Chapter 2
3, 5	horizontal	2, 2 digit	100 < sum < 200 # digits in sum ># digits in addends	Chapter 3
4	vertical	2, 3 digit	sums < 10	Chapter 3
6	horizontal	2, 4 digit	addition facts to 9	Chapter 2
7, 10, 12	vertical	5, vary from 2-4 digits	addition facts; place value; regrouping	Chapter 4
8	horizontal	3, 2 digit	addition facts; regrouping; using 0 in unit's place	Chapter 3
. . . . n				

The following is an example of a listed item format and how it may be covered in the curriculum:*

Item Format: 3 addends, sum < 10, horizontal
Curriculum Content:

0
00 00 0

These are John's. These are Jeff's. This is Karl's.

How many are there in all?
This is the number sentence we would use:

2 + 3 + 1 + _____

This material was found in Chapter 2, along with other problems of a similar type. Therefore, next to the item format, write "Chapter 2." At the end of Chapter 2, note that 5 test items were covered in that chapter. If the test contains 100 items, this would be 5% of the total items:

Unit	Items Covered	% of Total
Chapter 2	5	5%

The percentage of overlap between a particular curriculum and an achievement test is estimated by the cumulative percentage that should be calculated by adding each new unit's percentage to the previous percentage (see Example 2, page B-9). The percentage of overlap must be calculated for each curriculum (reading and mathematics) for each classroom. One number must be recorded for each curriculum for each classroom according to the following rules. Example 3 (page B-9) is provided to aid you in these calculations.

1. If the entire class is using the same single series and the entire class ends at a single spot in the curriculum, use the cumulative percentage of overlap of that spot.

---

*When assuming that a test item or test item content is covered in a text series, care should be taken to ensure that the manner of presentation is similar with respect to the availability of cues for the item's solution. For example, in the item set given above, if the student were always exposed to both the pictures and the equation, but the test only gave the equation, one could not assume the information had been taught.

Example 2

Unit	Items Covered	% of Total	Cumulative %
Chapter 1	6	6	6
Chapter 2	5	5	11
Chapter 3	9	9	20
.			
.			
.			
.			
.			
n			

Example 3

Name of Student	Curriculum Series Used	Final Curriculum Location (Page Number, Unit, Level, Skill Number, etc.)
Aloysius Alhazen	Abracadabra	48, J, 10
.		
.		
.		
.		
.		
n		

2. If the entire class is using the same single series but different students complete differing amounts of the curriculum, estimate the percentage of overlap for each student and average over all students. (Do not average groups of students.)
3. If the class is using different series (or books) for different groups of individuals, proceed as in (2).
4. If the class is using several series for each child, estimate the maximum percentage of overlap for each child, then average over all children.

3-4. In asking you to respond to these questions, we are trying to elicit your judgment about how motivating the curricula appear to be. Your familiarity with many curricula in both reading and mathematics should aid you in making a decision about how interesting students would find a particular curriculum. We don't expect that you have practical information (as does a teacher) about how much the children are interested in or intrigued by the material contained in the text. We want you to judge the motivational qualities of the materials from a strictly theoretical point of view, i.e., do the materials appear to be intrinsically interesting? Please rate the curriculum using your own experience and considering such things as pacing, format, artistry, etc.

5-6. For these questions, we are asking you to make a similar judgment. However, for the ratings, consider the main mode(s) of presentation such as textbook or workbook, and consider the variation that occurs within that mode. Consider variation in patterns of presentation, layout, and type of response required by the student. For example, a program that presents text followed by questions for the majority of instruction has little variation and would be rated 1. On the other hand, a program that has text followed by a variety of activities, such as writing short answers, writing plays, and writing questions for other students to answer, would have a large variation within that mode and would be rated 5.

7. A general objective is a global, program goal.

Examples of general reading objectives:

1. The child will become an independent reader, choosing to read for pleasure.
2. The child will be able to extract meaning from written text.

A specific objective is one that is linked with the instructional materials and is most often stated in behavioral terms.

Examples of specific reading objectives:

1. The child will be able to decode all one-syllable words in his/her reading text.
2. The child will be able to read a short paragraph story (three or four sentences) and paraphrase the story.

8. If no specific objectives are present in the reading curriculum, skip to Item 9. If there are some specific objectives in reading, answer this question:

Consider the clarity of the specific objectives in the reading curriculum. On a scale from 1 to 5, rate the clarity of the specific objectives.

A clear objective is one that is focused and narrow. It precisely states what the child will know how to do on completion of the materials. An example of a specific objective that is clear (i.e., rated 5) is: upon completing this unit, the student will be able to pronounce all one-syllable words of the CVC form.

An unclear objective is one that is dispersed and broad. It does not describe precisely what the child will be able to do on completion of the materials. An example of a specific objective that is unclear (i.e., rated 1) is: upon completing this unit, the child knows how to read one-syllable words.

9. A written statement of a reading objective may appear either in the teacher's manual or the child's text or in both. An example of a written statement of a reading objective is: the child will be able to pronounce any one-syllable words with two vowels, the final vowel being a silent "e."

An example of the new objective is: to present the child with a list of words that conform to this rule (sale, bale, male) and to pronounce the words for him/her.

A range of examples covers the broadest extremes of the domain. An example is: sale, Joe, ate.

Practice problems for the objective could be a list of words conforming to the rule that the child must pronounce by himself/herself.

A non-example for the objective is one outside of the domain. An example is: rut, not, near.

10. The materials match closely (5) if the contents of both the materials and the objectives are identical. The materials do not match the objectives (1) if the contents are not identical.

11. A placement procedure is a systematic method (either overall placement or unit pretest) built into the curriculum for assessing student initial abilities. It may be a diagnostic test, a student-teacher interview, a conference with the previous year's teacher, etc.

A monitoring procedure is a way (e.g., test, conference) of diagnosing which new material is prescribed. The procedure takes place prior to any attempt to teach materials.

A method for assessing mastery monitors how well a section of material has been learned. It may be a test, conference, etc., and is given after the child has studied the material.

If the reading curriculum has any of these procedures, please specify the method, such as test, conference, etc.

12. The reading curriculum and the student would be easy to match if there are a placement procedure built into the curriculum, a monitoring procedure for prescribing new materials, and a method for assessing mastery of studied material.

It would be difficult to match the reading curriculum and the student if all three of the above procedure were missing.

13. This question is very much like the overlap question (Item 2). It is concerned with the degree to which the placement tests that are built into the reading curriculum contain the same type of information as the curriculum. Unlike the overlap question, however, this question does not ask that you estimate the degree of overlap through a specific procedure. We ask that you make a judgment on the basis of your familiarity with the curriculum you are evaluating. No relation (1) between placement tests and the curriculum content means: (1) there are no placement tests built into the curriculum, or (2) the placement tests are in no way related to the curriculum content (e.g., a test for gross motor skills is used to place children in a reading group). High relation (5) means that for each item on the placement test there is a matching item in the curriculum content.

14. This question is the same type of question as Item 13. You are asked to judge the degree of overlap between the curriculum content in reading and mastery tests.

15. A general objective is a global, program goal.

Examples of general mathematics objectives:

1. The child will demonstrate problem-solving skills.
2. The child will enjoy playing mathematical games.

A specific objective is one that is linked with the instructional materials and is most often stated in behavioral terms.

Examples of specific mathematics objectives:

1. The child will be able to add two one-digit numbers.
2. The child will be able to count to 100 by fives.

16. If no specific objectives are present in the mathematics curriculum, skip to Item 17. If there are some specific objectives in mathematics, answer this question:

As you did for reading, consider for mathematics the clarity of the specific objectives. On a scale from 1 to 5, rate the clarity of the specific objectives.

A clear specific objective that would be rated 5 might be: given a series of one-digit two-numeral addition problems in the vertical form, the child will be able to correctly calculate their sums.

An unclear specific objective (1) might be: given a set of one-digit addition problems, the child will understand how to find their sums.

17. A written statement of a mathematics objective may appear either in the teacher's manual or the child's text or in both. An example of a written statement of a mathematics objective is: the student will be able to sum any two one-digit whole numbers.

An example of the new objective is: to present the child with a list of solved problems and conceptual methods for calculating correct answers to the problems (e.g., number line representations of the solutions to 3+1=4, 4+1=5, 1+2=3).

A range of examples covers the broadest extremes of the domain. An example is: 5+4=9, 8+9=17, 6+7=13.

Practice problems for the objective could simply be a list of unsolved problems for the student to complete.

A non-example for the objective is one that is outside of the domain. An example is: 1-1=?

18. The materials match closely (5) if the contents of both the materials and the objectives are identical. The materials do not match the objectives (1) if the contents are not identical.

19. A placement procedure is a systematic method built into the curriculum for assessing student initial abilities. It may be a diagnostic test, a student-teacher interview, a conference with the previous year's teacher, etc.

A monitoring procedure is a way (e.g., test, conference) of diagnosing which new material is prescribed. The procedure takes place prior to any attempt to teach materials.

A method for assessing mastery monitors how well a section of material has been learned. It may be a test, conference, etc., and is given after the child has studied the material.

If the mathematics curriculum has any of these procedures, please specify the method, such as test, conference, etc.

20. The mathematics curriculum and the student would be easy to match if there are a placement procedure built into the curriculum, a monitoring procedure for prescribing new materials, and a method for assessing mastery of studied material.

It would be difficult to match the mathematics curriculum and the student if all three of the above procedure were missing.

21. This question is very much like the overlap question (Item 1). It is concerned with the degree to which the placement tests that are built into the mathematics curriculum contain the same type of information as the curriculum. Unlike the overlap question, however, this question does not ask that you estimate the degree of overlap through a specific procedure. We ask that you make a judgment on the basis of your familiarity with the curriculum you are evaluating. No relation (1) between placement tests and the curriculum content means: (1) there are no placement tests built into the curriculum, or (2) the placement tests are in no way related to the curriculum (e.g., a test of color recognition is used to place children in mathematics). High relation (5) means that for each item on the placement test there is a matching item in the curriculum content.

22. This question is the same type of question as Item 21. You are asked to judge the degree of overlap between the curriculum content in mathematics and mastery tests.

23a-23b. A systematic method for assessing student initial abilities must be built into the curriculum for this question to be answered "yes." It may be a placement test, a student-teacher interview, a conference with the previous year's teacher, etc. It can be either an overall placement assessment or an assessment at the beginning of each unit. If no such method exists, answer "no."

If a specific method is present, please specify it.

24a-24b. A systematic method for assessing student mastery must be built into the curriculum for this question to be answered "yes." It may be a test, conference between student and teacher, etc. The assessment of what the student has learned must be made after he/she has studied the material. If no such method exists, answer "no."

If a specific method is present, please specify it.

25-26. It is very clear (5) what the next unit should be if the curriculum provides specific sequencing information. For example, the curriculum might list the following sequence: If a student(s) passes Unit 1, level C, he/she should go to Unit 1, level D. If a student(s) does not successfully master Unit 1, level C, he/she should go to Unit 1, level C alternate.

On the other hand, the curriculum may provide little or no sequencing information but the way in which the units of material are lettered or numbered in itself may provide clear information. For example, it may be obvious that after Unit 1 is completed, Unit 2 is the unit that should be covered next. In this case, the rating would also be 5.

It is unclear what the next unit should be (i.e., rated 1) if the curriculum does not provide adequate sequencing information. For example, at the end of Unit 1, level C, the curriculum may state: If a student(s) has not mastered the skills in this unit, review those skills before continuing on.

It is also unclear what the next unit should be if units may be covered in any sequence and no guidance is provided.

27-28. Check those items that correspond to the suggestions in the text.

29-30. An unordered sequencing system is one in which students proceed through units in any order.

A linear sequencing system is one in which there is a specified progression of units. No choice is given for varying this progression.

A branched sequencing system is one in which there are alternative paths available to students after completing each unit. When an alternative unit is chosen, there exists a new series of alternative units through which to progress.

There are more alternative routes in a modular system than in a branched sequencing system.

Curriculum Analysis Questionnaire:

Directions for Coding

1-2. Record the percentage of overlap (cumulative percentage) for each classroom according to the rules that were stated on the "Directions for Obtaining Measures."

3-6. Record the number of the rating directly.

7. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.
Thus, the measure consists of 4 binary codes, e.g., 01000 = only specific objectives exist.

8. Record the number of the rating directly.

9. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.

10. Record the number of the rating directly.

11. To code the first column, assign zero (0) to lines without checks and assign one (1) to lines with checks.

To code the "specific method" column, several steps are required. After the questionnaires are completed, make a list of all unique answers in this column. Many of these unique answers will be similar and can be clustered into the same category. This list of unique methods should then be used by one coder to create a master list of no more than five major categories of response. Using this master list, code the answers in the method column using 1 (yes) and 0 (no) for each category. Example: Suppose the unique answers in this column cluster into only three major categories: testing, conference with student, and discussion with teacher of previous school year. If the specific method for placement in a particular questionnaire is testing, then the code would be 100.

Follow this procedure for placement, monitoring, and mastery separately.

12-14. Record the number of the rating directly.

15. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.

16. Record the number of the rating directly.

17. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.

18. Record the number of the rating directly.

19. To code the first column, assign zero (0) to lines without checks and assign one (1) to lines with checks.

To code the "specific method" column, several steps are required. After the questionnaires are completed, make a list of all unique answers in this column. Many of the unique answers will be similar and can be clustered into the same category. This list of unique methods should then be used by one coder to create a master list of no more than five major categories of response. Using this master list, code the answers in the method column using 1 (yes) or 0 (no) for each category. Example: Suppose the unique answers in this column cluster into only three major categories: testing, conference with student, and discussion with teacher of previous school year. If the specific method for placement in a particular questionnaire is testing, then the code would be 100.

Follow this procedure for placement, monitoring, and mastery separately.

20-22. Record the number of the rating directly.

23a-24a. For each curriculum (mathematics and reading), record:

no = 0
yes = 1

23b-24b. After the questionnaires are completed, one coder should make a master list of all unique answers. Use the master list to create no more than five major categories of responses. Then code each answer using 1 (yes) and 0 (no). Do this for each curriculum.

25-26. Record the number of the rating directly.

27-30. Assign zero (0) to lines without checks.
Assign one (1) to lines with checks.

Analyses of the Reading and Mathematics Tests of the CTBS (Expanded Edition), Grade 1, Level B, Form S*

Letter Sounds, Test 1, pages 2-3

Format--Students must choose one of three letter(s) from an oral presentation and given a pictured referent.

Item	Position	Letter Sound	Unit of Material
1	end	m	
2	end	n	
3	end	g	
4	end	f	
5	middle	a	
6	middle	o	
7	middle	e	
8	middle	i	
9	middle	o	
10	middle	i	
11	middle	e	
12	middle	u	

---

*These analyses are to be used in determining the curriculum overlap measure (Items 1 and 2 in the Curriculum Analysis Questionnaire).

Letter Sounds (continued)

Item	Position	Letter Sound	Unit of Material
13	beginning	ch	
14	beginning	sh	
15	beginning	ch	
16	beginning	sh	
17	beginning	th	
18	beginning	bl	
19	beginning	gl	
20	beginning	tr	
21	beginning	dr	
22	beginning	br	

Word Recognition I, Test 2, pages 4-5

Format--Students must choose correct word from four choices given two oral presentations.

Item	Word	Unit of Material
1	tree	
2	girl	
3	sleep	
4	hand	
5	door	
6	cake	
7	clock	
8	rain	
9	truck	
10	kitten	
11	puppy	
12	money	
13	city	
14	animal	
15	sister	
16	children	
17	breakfast	
18	dollar	
19	head	

Reading Comprehension, Test 3, pages 6-10

Format--Students must read sentence(s) and select one of three pictures that correctly represents the sentence(s).

Item	Sentence(s) (Keyword(s) are underlined)	Unit of Material
1	Two girls can paint.	
2	He can open the window by himself.	
3	The girl is little. The box is big.	
4	It is night.	
5	Susan sees her birthday cake. It is on the table.	
6	Bill and Bob like to read.	
7	The children like to jump rope.	
8	Mother made him a cake. It was a surprise.	
9	This is a train. It is a toy.	
10	I can get down. Father will help Mother down.	
11	This is one book. It is in a box.	
12	These are cans. They are open.	
13	The bug is black. It is on the bed.	
14	Billy will dress like a clown for the party. Joan will dress like a rabbit.	
15	Mary has a flower.	
16	She let the kitten out.	
17	The prince took a drink and changed into a frog.	
18	The eggs are on the plate.	
19	This is a pet.	

Reading Comprehension (continued)

Item	Sentence(s) (Keyword(s) are underlined)	Unit of Material
20	See the finger. A fly is on it.	
21	Here is a boy, some people, and a pig.	
22	The woman has a dollar.	
23	Mrs. Brown is in the city with a balloon.	
24	Look at the big toy.	

Word Recognition II, Test 4, pages 11-12

Format--Students must choose one of four words that matches a pictured referent.

Item	Word	Unit of Material
1	cake	
2	girl	
3	clock	
4	hand	
5	door	
6	kitten	
7	sat	
8	hot	
9	apple	
10	rain	
11	coat	
12	street	
13	city	
14	truck	
15	kitchen	
16	money	
17	children	
18	dollar	
19	sisters	

Language I, Test 5, pages 13-15

Format--Students must listen to a sentence, then select one of four pictures that matches the sentence.

Item	Sentence (Keyword(s) are underlined)	Unit of Material
1	The boys are walking the dogs.	
2	The truck has crossed the bridge.	
3	The car pulls the truck.	
4	The boy will dig a hole soon.	
5	The glasses are on the table.	
6	The girl will jump into the water.	
7	The fish is watching the turtle.	
8	The boy that the girl chased fell down.	
9	The boy is followed by the dog.	

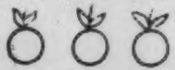
Format--Students must listen to four sentences, each presented twice, and select the one that does not violate standard language usage. The sequence of the sentences corresponds to the sequence of four pictures that are intended to serve as markers for each sentence.

Item	Part of Speech	Target Word(s)	Unit of Material
10	verb form	run	
11	verb form	fly	
12	plural noun		
13	verb form	buy	
14	plural noun		
15	comparative adjective	big, bigger, biggest	
16	comparative adjective	good, better, best	

Mathematics Concepts and Applications, Test 7, pages 18-21

Format--Students must choose one of three representations that correctly answers an oral question.

Examples of types of representation:

Type of Representation	Example
numerical	71
equation	12 - 3
series	46, 47, 48, 49
pictorial	
word	seven
algebraic equation	4 + □ = 10

Item	Type of Representation	No. Digits	Skill(s) Required	Unit of Material
1	numerical	2 digits	concept--smallest number	
2	equation	1 digit	add facts = 6	
3	series	2 digit	completing a series; pattern = +1	
4	numerical	2 digit	concept--largest number	
5	pictorial	----	ordinal numbers; concept--fifth	
6	pictorial and numerical	1 and 2 digit	match number of things to numeral	
7	equation	2 digit	tens and ones; rename numeral	

Mathematics Concepts and Applications (continued)

Item	Type of Representation	No. Digits	Skill(s) Required	Unit of Material
8	series	1 digit	counting by twos	
9	pictorial and word	1 digit	match number of things to word label	
10	series	2 digit	identify series with pattern = +1	
11	equation	2 digit - 1 digit	subtract facts = 9	
12	pictorial	1 digit	identify representation of addition word problem	
13	pictorial	1 digit	identify representation of subtraction word problem	
14	equation	1 digit	identify correct number sentence; add and subtract facts	
15	numerical	1 digit	solve word problem; add facts < 10	
16, 18	algebraic equation	1 digit	given missing numeral, find correct addition equation	
17	equation	1 digit	identify number sentence for addition word problem	
19, 20	pictorial	----	identify time on a clock face at half-hour	

Mathematics Concepts and Applications (continued)

Item	Type of Representation	No. Digits	Skill(s) Required	Unit of Material
21	pictorial	----	identify measurement of one inch with ruler equivalent	
22	pictorial	----	coin equivalents of 15¢	
23	pictorial	----	coin equivalents of 20¢	
24	pictorial	----	coin equivalents of a "quarter"	

Mathematics Computation, Test 8, pages 22-23

Format--Students must solve an addition problem and choose the correct response from three choices.

Item	Format	No. Digits and No. Numerals	Skill(s) Required	Unit of Material
1, 2, 4,	vertical	2 digit + 1 digit	add facts < 10	
3	vertical	2 digit + 2 digit	using 10 as addend	
			add facts < 10	
5, 6, 7	vertical	2 digit + 2 digit	add facts < 10	
8, 9, 11	vertical	4 1-digit numerals	add facts > 10	
10	vertical	4 1-digit numerals	using 0 as addend	
			add facts > 10	
12	horizontal	1 digit + 2 digit	regrouping	
			add facts = 10	
13, 16	horizontal	2 digit + 1 digit	using 10 as addend	
			add facts < 10	
14	horizontal	2 digit + 2 digit	add facts > 10	
15	horizontal	1 digit + 1 digit	add facts < 10	

Format--Students must solve a subtraction problem and choose the correct response from three choices.

Item	Format	No. Digits	Skill(s) Required	Unit of Material
17	vertical	1 digit - 1 digit	subtraction facts < 10	
18	vertical	2 digit - 1 digit	using 0 as subtracted number	

Mathematics Computation (continued)

Item	Format	No. Digits	Skill(s) Required	Unit of Material
19, 22	vertical	2 digit - 1 digit	subtraction facts < 10; answer has 0 in one's place	
20, 26	vertical	2 digit - 2 digit	subtraction facts < 10; answer has 0 in one's place	
21, 23, 25	vertical	2 digit - 1 digit	subtraction facts < 10	
24	vertical	2 digit - 2 digit	subtraction facts using tens	
27	vertical	2 digit - 2 digit	subtraction facts < 10	
28, 29	horizontal	2 digit - 1 digit	subtraction facts < 10; 1 as subtracted number	
30	horizontal	2 digit - 1 digit	subtraction facts > 10	
31	horizontal	2 digit - 1 digit	subtraction facts < 10	
32	horizontal	2 digit - 1 digit	subtraction facts > 10; regrouping	

Analyses of the Reading and Mathematics Tests of the CTBS (Expanded Edition), Grade 3, Level 1, Form T, Book 1*

Reading Vocabulary, Test 1, pages 2-4

Format--Words that mean the same (in phrase context).

Item		Unit of Material
1	leaping - jumping	
2	hind - back	
3	above - over	
4	sniff - smell	
5	saving - keeping	
6	outfit - costume	
7	become - grow	
8	chore - job	
9	harm - hurt	
10	customer - buyer	
11	chip - piece	
12	perhaps - maybe	
13	trail - path	

---

*These analyses are to be used in determining the curriculum overlap measure (Items 1 and 2 in the Curriculum Analysis Questionnaire).

Reading Vocabulary (continued)

Item		Unit of Material
14	vanishing - disappearing	
15	drag - pull	
16	hope - wish	
17	aid - help	
18	pardon - forgive	
19	again - over	
20	pound - beat	
21	obey - follow	
22	trust - belief	
23	guest - visitor	
24	voyage - journey	
25	scowling - frowning	
26	marsh - swamp	
27	protect - guard	
28	meadow - grassland	
29	accept - take	
30	usually - mostly	
31	principal - main	
32	tug - yank	
33	direct - guide	
34	upset - overturn	
35	cause - reason	
36	ruin - spoil	
37	rules - directions	
38	intend - plan	
39	anxious - worried	
40	encourage - support	

Reading Comprehension, Test 2, pages 5-10*

Format--Written passages followed by a set of items related to that passage.

Passage	Type	No. Lines
1	narrative with dialogue	22

Item	Format	Levels of Comprehension Skills Required	Unit of Material
1	question	inferential--supporting detail (one-word answer)	
2	question	literal--main idea (semantic substitute, sentence answer)	
3	question	inferential--cause and effect (sentence answer)	
4	question	inferential-sequence (sentence answer)	
5	question	literal--supporting detail (sentence answer)	

---

*The levels of comprehension skills required correspond in large part to those of Thomas C. Barrett in "Taxonomy of Cognitive and Affective Dimensions of Reading Comprehension," a chapter in Innovation and Change in Reading Instruction, edited by H. M. Robinson (Chicago: University of Chicago Press, 1968).

Reading Comprehension (continued)

Item	Format	Levels of Comprehension Skills Required	Unit of Material
6	completion	vocabulary-idiomatic expression "to make up her mind"--"to decide" (infinitive verb answer)	
7	question	inferential--emotional response of character (one-word answer)	
8	question	literal-sequence (semantic substitute, phrase answer)	
9	question	reorganization--synthesizing (phrase answer)	

Passage	Type	No. Lines
2	narrative with dialogue	33

Item	Format	Levels of Comprehension Skills Required	Unit of Material
10	question	inferential incidental detail (phrase answer)	
11	completion	reorganization synthesis (phrase answer)	
12	question	inferential character trait (one-word answer)	

Reading Comprehension (continued)

Item	Format	Levels of Comprehension Skills Required	Unit of Material
13	question	inferential incidental detail (phrase answer)	
14	question	inferential main idea (sentence answer)	
15	completion	inferential main idea choose best title (phrase answer)	
16	completion	vocabulary figured-guessed (one-word answer)	
17	question	inferential--narrator (one or two words, names, for answer)	
18	question	inferential emotional response of character (one-word answer)	

Passage	Type	No. Lines
3	narrative with dialogue	33

Item	Format	Levels of Comprehension Skills Required	Unit of Material
19	question	literal supporting detail (phrase answer)	
20	question	literal incidental detail (phrase answer)	

Reading Comprehension (continued)

Item	Format	Levels of Comprehension Skills Required	Unit of Material
21	question	inferential main idea (phrase answer)	
22	question	inferential sequence (sentence answer)	
23	question	reorganization synthesis (phrase answer)	
24	question	literal cause and effect (sentence answer)	
25	question	inferential transformation (one-word answer) (Maybe blue bikes are just slower--red bikes might go faster)	
26	completion	vocabulary bet--claim (one-word answer)	
27	question	inferential main idea (phrase answer)	
28	question	inferential emotional response of character (one-word answer)	

Reading Comprehension (continued)

Passage	Type	No. Lines
4	historical narrative	32

Item	Format	Levels of Comprehension Skills Required	Unit of Material
29	question	inferential emotional response of character (one-word answer)	
30	question	literal incidental detail (phrase answer)	
31	completion	literal incidental detail (semantic substitute, phrase answer)	
32	question	literal supporting detail (semantic substitute, sentence answer)	
33	completion	inferential supporting detail (phrase answer)	
34	completion	reorganization synthesis (phrase answer)	
35	completion	reorganization synthesis (phrase answer)	

Reading Comprehension (continued)

Item	Format	Levels of Comprehension Skills Required	Unit of Material
36	question	literal incidental detail (phrase answer)	
37	question	inferential main idea (moral) (sentence answer)	
38	question	inferential supporting detail (one-word answer)	

Passage	Type	No. Lines
5	poem	16

Item	Format	Levels of Comprehension Skills Required	Unit of Material
39	question	literal single-word vocabulary (phrase answer)	
40	completion	inferential/interpreting figurative language (one-word answer)	
41	question	inferential antecedent from "we" (one-word and phrase answer)	
42	question	reorganization, classification (one-word answer)	

Reading Comprehension (continued)

Item	Format	Levels of Comprehension Skills Required	Unit of Material
43	question	vocabulary sly--clever (one-word answer)	
44	question	inferential main idea choose best title (phrase answer)	
45	question	inferential antecedent from "we" (one-word answer)	

Language Expression, Test 5, pages 13-15

Format--A sentence is presented that is graphically segmented into three parts, with a fourth answer possible. Students must locate the segment containing a syntactic anomaly if one occurs.

Item	Location	Skill	Unit of Material
21	none	recognize no error; e.g., sentence grammatically correct	
22	verb phrase	locate error in use of plural form; e.g., ways--way	
23	verb phrase	locate error in use of verb; e.g., learned--taught	
24	verb phrase	locate inappropriate use of adverb; e.g., "most best"	

Format--Sentences are presented with a missing word that the student must select from one of four choices.

Item	Location of Blank	Skill	Unit of Material
25	middle	choose appropriate verb form (eat)	
26	middle	choose appropriate pronoun (who)	
27	end	choose appropriate pronoun (himself)	
28	beginning	choose appropriate verb form (were)	

Language Expression (continued)

Item	Location of Blank	Skill	Unit of Material
29	middle	choose appropriate verb form (taken)	
30	middle	choose appropriate verb form (has)	

Format--A four-line poem is presented with three blanks in lines 1-3. The student must select an appropriate word to fill in the blank or find a synonym.

Item	Location of Blank	Skill	Unit of Material
31	end of line	choose appropriate verb (grow)	
32	end of line	choose appropriate object (trees)	
33	beginning of line	choose appropriate pronoun (they)	
34	none	vocabulary--one word (breeze--wind)	

Language Expression (continued)

Format--Given one word and four sentences using that word, students must find the incorrect usage where graphic similarity is a distractor.

Item	Keyword	Content	Unit of Material
35	ear	noun used as verb e. g., ear--hear	
36	color	incorrect adjective e. g., color--clear	
37	drink	incorrect object e. g., drink--brink	
38	rope	incorrect verb e. g., rope--rode	

Format--A sentence is presented with an underlined word that has multiple meanings. Students must choose the correct meaning using sentence context.

Item	Keyword	Response	Unit of Material
39	dip	scoop	
40	tracks	footprints	
41	fair	sunny and clear	
42	base	place on a baseball field	

Language Expression (continued)

Item	Format	Skill	Unit of Material
43	question	vocabulary transformation not shine--smoke	
44	question	vocabulary transformation most sticky--glue	
45	question	vocabulary transformation not grow--pillow	

Format--Given a four-sentence story in scrambled order, students must choose the most appropriate first sentence.

Item	Content	Unit of Material
46	general to specific	
47	temporal sequencing	
48	general to specific	
49	temporal sequencing	
50	temporal sequencing	

Mathematics Computation, Test 6, pages 16-17

Examples of format:

Addition:

Horizontal: $A + B =$

Vertical: $+\begin{array}{c} A \\ \underline{B} \end{array}$

Subtraction:

Horizontal: $A - B =$

Vertical: $-\begin{array}{c} A \\ \underline{B} \end{array}$

Multiplication:

Horizontal: $A \times B =$

Vertical: $\times\begin{array}{c} A \\ \underline{B} \end{array}$

Division:

Vertical: $B\overline{)A}$

Fraction: $\frac{A}{B}$

Mathematics Computation (continued)

Item	Format	No. Numerals and No. Digits	Skill(s) Required	Unit of Material
1	vertical	3, 1 digit	$10 < \text{sum} < 20$, using 0 as an addend	
2	horizontal	3, 1 digit	$10 < \text{sum} < 20$	
3,5	horizontal	2, 2 digit	$100 < \text{sum} < 200$ # digits in sum > # digits in addends	
4	vertical	2, 3 digit	sums $\gtrless 10$	
6	horizontal	2, 4 digit	addition facts to 9	
7, 10, 12	vertical	5, vary from 2-4 digits	addition facts; place value; regrouping	
8	horizontal	3, 2 digit	addition facts; regrouping, using 0 in unit's place	
9, 11	vertical	4, vary from 1-2 digits	addition facts; place value; regrouping	
13	vertical	3 digit - 3 digit	subtraction facts < 10, using 0 in unit's place in both numerals	
14	horizontal	2 digit - 1 digit	subtraction facts < 10; place value	
15	horizontal	2 digit - 2 digit	subtraction facts < 10, using 0 in unit's place of subtracted numeral	
16	horizontal	2 digit - 1 digit	subtraction facts; regrouping; place value	

Mathematics Computation (continued)

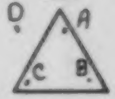
Item	Format	No. Numerals and No. Digits	Skill(s) Required	Unit of Material
17	vertical	2 digit - 2 digit	subtraction facts < 10; 1-digit answer	
18	horizontal	3 digit - 3 digit	subtraction facts < 10, using 0 in ten's place of subtracted numeral	
19	vertical	3 digit - 2 digit	subtraction facts; regrouping	
20	horizontal	3 digit - 2 digit	subtraction facts < 10; 3-digit answer	
21	horizontal	3 digit - 2 digit	subtraction facts; regrouping; 2-digit answer	
22	vertical	3 digit - 1 digit	subtraction facts ≤ 10; hundreds as numeral subtracted from; regrouping	
23	horizontal	3 digit - 2 digit	subtraction of 10 from hundreds; regrouping	
24	vertical	3 digit - 3 digit	subtraction facts; regrouping; 1-digit answer	
25	vertical	2 digit x 1 digit	multiplication facts using 10	
26	horizontal	2 digit x 1 digit	multiplication facts using 20	
27	vertical	1 digit x 1 digit	multiplication facts using 0	

Mathematics Computation (continued)

Item	Format	No. Numerals and No. Digits	Skill(s) Required	Unit of Material
28, 31	horizontal	1 digit x 1 digit	multiplication facts	
29, 34	vertical	3 digit x 1 digit	multiplication facts	
30	vertical	2 digit x 1 digit	multiplication facts; digits in product > digits in factors	
32, 35	vertical	2 digit x 1 digit	multiplication facts; digits in product > digits in factors; regrouping	
33	horizontal	1 digit x 2 digit	multiplication facts; digits in product > digits in factors	
36	vertical	3 digit x 1 digit	multiplication facts; regrouping	
37	vertical	1 digit ÷ 1 digit	division facts by 1	
38, 43	horizontal	1 digit ÷ 1 digit	division facts	
39, 44	fraction	1 digit ÷ 1 digit	division facts	
40	vertical	3 digit ÷ 2 digit	division of 100 by 10	
41	fraction	2 digit ÷ 1 digit	division facts by 2, using tens as dividend	
42, 46	vertical	1 digit ÷ 1 digit	division facts by 6	
45, 48	vertical	2 digit ÷ 1 digit	division facts; 2-digit quotient	
47	vertical	2 digit ÷ 2 digit	division facts; quotient = 1	

Mathematics Concepts and Applications, Test 7, pages 18-24

Explanations and examples of format:

Format	Explanation	Example
question	refers to the fact that students must read all or part of a written question in order to respond	Which tree below is in the middle of the row?
word problem	refers to the traditional interpretation; in order words, the problem is embedded in verbal context only	Kimi had 11 apples. She gave 6 apples away and ate 2. How many apples did she have left?
visual	refers to the presence of a visual representation that is needed to respond to the question	
geometric form	refers to the presence of geometric shapes that are needed to respond to the question	
algebraic form	refers to the presence of an equation with a missing numeral that uses a letter or box to represent the unknown	13 - n = 4

Mathematics Concepts and Applications (continued)

Item	Format	Skill(s)	Unit of Material
1	question--1 keyword geometric forms	match verbal label (triangle) to drawn shape	
2	question--1 keyword visual	discriminate one from a series of identical shapes; concept of middle in a series with no size or shape clue	
3	question--all visual	concept of equal to	
4	question--key phrase geometric forms	rotating angles and match to sample	
5	question--1 keyword visual with labels in response	measurement with ruler equivalent	
6, 10	question	equivalent expressions (e.g., another name for 11 is 12 - 1)	
7	addition word problem	add facts < 10 "set" as distractor in stem	
8	question--first 3 words as keywords--visual	measure with liter equivalent	
9	question--visual	Venn diagram; how many in 1 of 3 intersecting sets	
11	question--visual response	sets; even numbers	
12, 17	question	place value	
13	word problem--1 keyword visual	measure time to half-hour; recognition of time notation, e.g., 2:30; read finish	

Mathematics Concepts and Applications (continued)

Item	Format	Skill(s)	Unit of Material
14	question--visual	fractional part; 1/4	
15	question	concept of "greater than" using a 5-digit numeral; large number of distractors	
16	question	common objects named; choose object whose shape is equivalent to a rectangle	
18, 22, 25	question--algebraic form	equivalent expressions	
19	question--geometric form	points inside and outside a triangle	
20	word problem	2 steps; identify operations; subtract and add	
21	question--geometric form	label line segments; identify diameter of a circle	
23	question	odd numbers	
24	question--visual	measure with grams to 100 units	
26	question--visual	how many in a set of shapes including 2 characteristics; 1 consistent (shape), 1 inconsistent (shading)	
27	question--visual	inequality of sets; how many more	
28, 35	completion--algebraic form	solve for missing numeral as addend	
29	word problem	2 steps; subtract twice or add and subtract; add and subtract facts	

Mathematics Concepts and Applications (continued)

Item	Format	Skill(s)	Unit of Material
30	word problem	time notation and clock time; subtract time using different forms of presentation; response as straight numeral	
31	question--visual	identify number of common elements from 2 sets and combine; add facts < 10	
32	completion--geometric form	diagonal line segment of square divides it into 2 triangles	
33	question--visual	find common elements from 2 sets; visual response	
34	question--visual	add facts; 4 numerals; follow map	
36	question--visual	add facts; 4 numerals; concept of "around" a square (or perimeter)	
37	word problem	read "o'clock" form of time; add hours to given time	
38	word problem	subtraction; 2 digit - 1 digit; regrouping	
39	word problem	measurement; convert quarts to cups; 1-digit multiplication facts by 4	
40	word problem	addition; 2 digits, 3 numerals; regrouping	
41	word problem	convert feet to inches; concept of half	
42, 44	question--algebraic form	equality of equations; 2 steps	

Mathematics Concepts and Applications (continued)

Item	Format	Skill(s)	Unit of Material
43, 45	word problem	subtraction; smaller number given first; 2 digit; regrouping	
46	word problem	multiplication facts by 5	
47	word problem	division by 5; has "divided" cue	
48	word problem	multiply by 2; concept twice as many	
49	word problem	division by 7; has "divided" cue	
50	word problem	multiply by 3; has "times" cue	

APPENDIX B-2

BEHAVIORAL ANALYSIS OF CURRICULA*

Page

Behavioral Analysis of Curricula Data Sheet....... B-53

Behavioral Analysis of Curricula Data Sheet: Instructions........................ B-57

Behavioral Analysis of Curricula Coding Sheet B-67

Additional Information...................... B-68

---

*The measures included in this Appendix are designed to assess the quality of individualized instructional materials in both mathematics and reading. The measures require the presence of diagnostic tests and procedures for individual decision making in instruction. Therefore, they cannot be estimated for all curricula in the study. It is recommended that they be estimated whenever feasible and that separate analyses be performed on classrooms for which these measures are obtained.

Behavioral Analysis of Curricula Data Sheet

Name of Curriculum Expert ____________________

Date ____________________

School District ____________________

School ____________________

Grade ____________________

Teacher's Name ____________________

Measures of Adapting Variables

1. Consequence ratio. Using the in-class interview, Items 68 and 69, identify the most recent diagnostic test taken by each child. Locate each test in your copy of the curriculum materials and identify an appropriate "unit" to be counted and compared across classrooms. This unit must be constant within classrooms. Count the number of units in each diagnostic test and add them. 1. __________

2. Using the teacher's edition of the curriculum material as your guide, locate all the material the developer suggests be prescribed when diagnostic test performance does not indicate mastery. Count the units in the developer's recommended consequent material for each diagnostic test. Enter this number. 2. __________

3. Using the in-class interview, Items 68 and 69, identify the teacher's prescription for each diagnostic test failure. Locate this prescription in your copy of the materials and compare, in terms of number of arbitrary units, the developer's recommended prescription. Note the amount and direction of each deviation, and enter a total correction number (with plus or minus sign) for the teacher's modification of the developer's prescription. 3. __________

4. Using the in-class interview, Items 68 and 69, check for any teacher prescription of material consequent to a diagnostic test pass. Since the developer ordinarily recommends that teaching material be skipped consequent to a pass, the teacher who prescribes material consequent to a pass is modifying the actual consequence time. Count the number of units in all material prescribed consequent to a pass and enter that number.

4. ______

5. To adjust the developer's recommended consequence for actual classroom use, add or subtract (depending on sign) the number entered on line 3 to/from the number entered on line 2.

5. ______

6. As a second adjustment of the developer's recommended consequence for actual classroom use, subtract the number of units entered on line 4 from the number entered on line 5.

6. ______

7. Add lines 1 and 6 to find the "time," i.e., total number of units, used in both diagnosing and teaching.

7. ______

8. The consequence ratio is consequence "time" divided by total "time." Divide the number entered on line 6 by the number entered on line 7. Take this division to two significant places and enter the number on line 11 and on the coding sheet (page B-67). This is the consequence ratio.

8. ______

9. Discriminability ratio. Using data from the same in-class interview question, count the number of students who passed a diagnostic test.

9. ______

10. Using data from the same in-class interview question, count the number of students who failed a diagnostic test.

10. ______

11. The discriminability ratio is the number of diagnostic test passes or failures (whichever is smaller) divided by the total number of students in the class. Select the smaller number of lines 9 and 10 and divide this smaller number by the number of students in the class, counting students who did not take a diagnostic test. Take this to two significant places.

11. ______

12. Double the number entered on line 11 and enter on line 12 and on the coding sheet. This is the discriminability ratio. 12. ____________

13. Predictive validity ratio.* Using data from the in-class interview, Items 68 and 69, count the number of students whose diagnostic test performance indicates a "skip" and who also passed the mastery test. Enter this number. 13. ____________

14. Using data from the in-class interview, Items 68 and 69, count the number of students whose diagnostic test performance indicated a "take" and who also failed the mastery test. Enter this number. 14. ____________

15. Add the entries for lines 13 and 14 to find the number of correct predictions. 15. ____________

16. Using data from the same in-class interview question, count the total number of students for whom a diagnostic test decision was made. Enter this number. 16. ____________

17. The predictive validity ratio is the number of correct predictions divided by the number of decisions. Divide the entry on line 15 by the entry on line 16. Take this division to two significant places and enter this number on line 17 and on the coding sheet. This is the predictive validity ratio. 17. ____________

18. Select the lowest number of lines 8, 12, and 17. Enter this lowest number on line 18 and on the coding sheet. 18. ____________

Measure of Appropriately Determined Response Opportunities

19. For each child's assignment in reading class, determine the number of response opportunities. Sum these individual totals to find a classroom total of response opportunities. Enter this classroom total. 19. ____________

---

[*] Because this measure requires classroom intervention that may be difficult for the contractor to accomplish, it should be considered an optional measure.

20. For each response opportunity, answer "yes" or "no" to the following question: "Can a child respond correctly without behaving in the manner described by the stated or inferred objective?" Sum the number of "yeses" to this question and enter this number. 20. ___________

21. For those response opportunities to which you answered "no" to the question on line 20, answer "yes" or "no" to the following question: "Is the child likely to answer incorrectly because the material responded to is unclear, or because previous material has not prepared the child to make such a response?" Sum the number of "yeses" to this question and enter this number. 21. ___________

22. Subtract the number of potentially inappropriately cued responses (20) and the number of potential errors (21) from the total number of classroom response opportunities (19). Enter this number of appropriately determined response opportunities per classroom. 22. ___________

23. Divide the number of appropriately determined response opportunities per classroom (22) by the number of children with reading assignments in the classroom. Enter this number on line 23 and on the coding sheet. 23. ___________

24. Divide the number of appropriately determined response opportunities (22) by the total number of response opportunities (19). Enter this percentage on line 24 and on the coding sheet. 24. ___________

Behavioral Analysis of Curricula Data Sheet: Instructions

Measures of Adapting Variables

1. Consequence ratio. Diagnostic testing should save students' time by ensuring that they only study material that they do not already know. However, since diagnostic testing comes at a cost in student time, it is necessary to have a measure of the cost-effectiveness of such testing. The consequence ratio, which relates time spent taking diagnostic tests to total time (i.e., diagnostic test time plus prescribed teaching time) is such a measure. To calculate each classroom's total diagnostic test "time," you will use information from the in-class interview.

   Using information from in-class interview Items 68 and 69, locate each diagnostic test that a child in the classroom has taken. If the test is in a published curriculum, the interview data provide title and page numbers. If the test has been developed by the teacher, a copy of it with subject identification number is attached to the interview questionnaire.

   Next, examine each diagnostic test to select a unit to stand for test "times." The word "time" is used because the most accurate consequence ratios are calculated by actually timing the students as they perform the diagnostic test and consequent material. However, almost as good an estimate can be made by selecting an artificial "unit" to stand for time. Since the result is a ratio, it is important that the unit counted in the diagnostic test material be identical to the unit counted in the consequent teaching material. After examining the diagnostic tests and teaching material, you can select an arbitrary unit that seems appropriate for the curriculum. For example, the number of math problems the student does in the diagnostic test and in the consequent material might be counted and compared. In a reading curriculum, the number of words in the diagnostic test might be compared to the number of words in the consequent material. As long as the unit counted in the diagnostic test and in the consequent material is identical, such things as number of pages, lines, etc., may be compared.

   Once the unit to be compared is selected, count the number of these units in each diagnostic test taken by each child in the classroom. You are interested in a classroom total of diagnostic test units or "time." For example, given that diagnostic test A, containing five units, had been taken by two children, and that diagnostic test B, containing three units, had been taken by one child, the classroom total of diagnostic test units would be 13 (5 units x 2 children + 3 units x 1 child). Enter this classroom number of diagnostic test units on the data sheet.

2. In calculating the consequence ratio, it is irrelevant whether the student actually took or skipped the material. The ratio simply identifies the cost-effectiveness of diagnostic testing as compared to the potential time spent in teaching and diagnosing. Therefore, each diagnostic test time included in the ratio must have an accompanying consequence time. Included in the consequence "time" of a diagnostic test are all readings, exercises, lessons, tests, etc., that a student who fails the diagnostic test must take and that a student who passes the diagnostic test may skip.

Usually in individualized curriculum materials, the consequent material for each diagnostic test is clearly identified in the teacher's manual. Thus, a consequence ratio could be calculated by examining the curriculum materials alone without reference to their use in a classroom. However, a teacher frequently will use the materials in ways that vary from the format identified in the teacher's manual. For example, if a teacher prescribes fewer assignments than the curriculum suggests consequent to performance on a diagnostic test, the actual classroom consequence ratio would differ from a ratio calculated by examining the materials alone.

Classroom information alone cannot be used to calculate a consequence ratio, since there is no in-class record of the consequence for children who "skip" material. Thus, we suggest that you begin by identifying the developer's recommended consequence for each diagnostic test taken, and then correcting this figure with the available classroom use information.

For each published diagnostic test recorded on in-class interview Items 68 and 69, locate the publisher's recommended consequence material. Using the same arbitrary unit counted in the diagnostic tests (line 1), estimate recommended consequence "time." Remember that we are concerned with the potential consequence of failing a diagnostic test. Thus, a consequence "time" must be entered for each diagnostic test taken by each child, even if some children in the class passed the diagnostic test and so avoided the consequence. For example, assume that diagnostic test A has 5 units with a potential consequence of 20 units. Mary passed and JoAnn failed test A. The total potential consequence time is 40 units.

For each teacher-developed diagnostic test (if any), locate material actually or potentially prescribed consequent to test performance. This material is also identified on the in-class interview, Items 68 and 69. Count the units to estimate this consequence "time."

Add the developer's recommended consequence "time" and the teacher-developed consequence "time" and enter this figure on line 2 of the data sheet.

3. To correct for teacher changes in developer recommendations, use data from the same in-class interview question. You can identify the actual prescriptions for children whose diagnostic test performance indicated the necessity for teaching material. Compare the actual prescription to the developer's recommended prescription in terms of number of units. If the prescriptions are identical for a child, note zero correction. If, for example, a child's actual prescription is 5 units shorter than the recommended prescription, note a correction of -5; if a child's actual prescription is 7 units longer than the recommended prescription, note a correction of +7. Add all corrections, attending to sign in each case. The above three examples give a total correction figure of +2. Enter your total correction figure and the appropriate sign on line 3 of the data sheet.

4. This correction figure is to take care of the circumstance in which a teacher gives diagnostic tests, yet assigns teaching material even to students whose test performance indicates a "skip." Although, ordinarily, the consequence ratio is calculated without reference to actual student performance, in this unusual case the teacher's modifications are actually at a cost in student time above a nonindividualized version of that curriculum in which the students simply take all the teaching material.

   To identify prescribing consequent to a "skip," note each student's score on the diagnostic test (in-class interview Items 68 and 69) and use the developer's recommendations to analyze that score as indicating a "skip" or "take." If the teacher has prescribed material that students' scores indicate may be skipped, count the number of units in this prescribed material and enter this number on line 4 of the data sheet.

5. Correct the number of units entered on line 2 by adding or subtracting (depending on sign) the correction figure entered on line 3. Enter this corrected consequence number on line 5.

6. Correct the number of units entered on line 5 by subtracting from it the number entered on line 4. Enter this as the final corrected consequence "time" on line 6 of the data sheet.

7. The consequence ratio is the consequence "time" divided by the sum of testing "time" and consequence "time." To find this sum, add lines 1 and 6.

8. Divide line 6 by line 7. Take this division to two significant places. Enter this number on line 8 of the data sheet and on the coding sheet.

9. Discriminability ratio. A diagnostic test should reflect individual differences. If all students pass or all students fail the diagnostic test, their prescriptions are identical. No individual differences have been detected and, thus, we might expect that no benefits of individualization would be detected.

Using data from the in-class interview, Items 68 and 69, count the number of students who passed a diagnostic test, including both published and teacher-developed tests. Enter this number on line 9.

10. Using data from the same in-class interview question, count the number of students who failed a diagnostic test, including both published and teacher-developed tests. Enter this number on line 10.

11. The discriminability ratio is the number of students who passed or failed (whichever is smaller) divided by the number of students in the class. Select the smaller number from lines 9 and 10. (If no students were given diagnostic tests, the "smaller" number is zero, and zero divided by the number of students in the class is still zero.) Count the number of students in the class, including those who did not take a diagnostic test. Divide this total number of students in the class into the selected smaller number. Take this division to two significant places and enter this number on line 11 of the data sheet.

12. The discriminability ratio as calculated in line 11 varies from 0 to .50. To make it more directly comparable to the consequence ratio and the predictive validity ratio, which vary from 0 to 1.0, it is simply doubled and entered on the coding sheet.

13. Predictive validity ratio. A valid diagnostic test predicts whether students need teaching material before being able to pass the post-test or mastery test. The diagnostic test and mastery test purport to measure the same student behaviors. Therefore, the diagnostic test should be a good predictor of performance on the mastery test if the student is not given intervening teaching material. If no mastery test corresponding to a diagnostic test can be identified in a particular curriculum, data derived from a second administration of the same diagnostic test can be treated exactly as if from a mastery test.

Attached to the in-class interview, Items 68 and 69, you will find data that have been gathered as follows. During a specified time period of approximately one month, the teacher was asked to identify the first diagnostic test taken by each child and record the child's performance on that test. Then prior to any attempt to teach those objectives, the teacher administered the corresponding mastery test, or readministered the same diagnostic test. On the data sheet, location of the mastery test is identified and the child's performance recorded.

Using a teacher's edition of the curriculum, check to see that each child's diagnostic test and mastery test do correspond as recommended by the curriculum developers. In other words, this is not a judgment about whether the tests indeed measure the same behavior, but rather a check on whether the teacher correctly identified the recommended mastery test. Data from children whose two tests do not "match" according to recommendations of the developer cannot be used to obtain a measure of diagnostic test predictive validity. Eliminate all such "mismatches" from your data sheet.

When a student's diagnostic test performance indicates that teaching material can be skipped, the prediction is that the student would pass the corresponding mastery test without any intervening teaching material. Thus, for students whose diagnostic test performance is a "skip" and who also pass the mastery test, the diagnostic test has correctly predicted mastery test performance. Count these students on the data sheet and enter this number on line 13.

14. When a student's diagnostic test performance indicates that teaching should be taken, the prediction is that the student would fail the mastery test unless he or she took some intervening teaching material. Thus, for students whose diagnostic test performance is a "take" and who also fail the corresponding mastery test, the diagnostic test has predicted correctly. Count these students on the data sheet and enter this number on line 14.

15. Add the entries for lines 13 and 14 to find the number of correct predictions. Enter this number.

16. If the diagnostic test indicates a "skip," yet the student fails the mastery test, or if the diagnostic test indicates a "take," yet the student passes the mastery test, the diagnostic test has predicted incorrectly. All decisions that are not correct predictions should fit in one of the above two categories. Using the data sheet, count the students for whom predictive decisions (skip or take) were made. (Do not include students whose tests were mismatched as explained in 13.) Enter this number of decisions on line 16.

17. The predictive validity ratio is the number of correct predictions divided by the number of decisions (correct plus incorrect predictions). Divide the entry for line 15 (correct predictions) by the entry for line 16 (total decisions). Take this division to two significant places and enter this number on line 17 of the data sheet and on the coding sheet.

If you have a data sheet on predictive validity for some but not all classrooms, you may enter the value for predictive validity derived from one classroom in the calculations for other classrooms using that same curriculum.

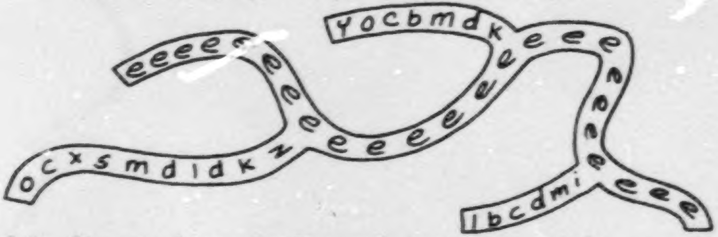
18. Each measure of the goodness of adapting can be correlated alone with end-of-year student achievement. However, direct relationships would not be expected since two programs could be equally high in discriminability but differ greatly on the other two measures. Yet, analyzing separately for each measure can provide information on how one factor contributes to overall achievement if other factors are held constant. It is also of interest, however, to get an overall estimate of the value of adapting. Since the three measures are of interrelated factors, an improvement in one measure comes at the cost of another. The lowest value of the three, then, represents the effectiveness of that adaptive feature more accurately than does any kind of average. Enter the lowest value of the three ratios on line 18 of the data sheet and on the coding sheet.

Measure of Appropriately Determined Response Opportunities

19. Information enabling you to locate each child's assignment should be taken from the in-class interview, Items 68 and 69. If different children are working on different assignments, be sure to locate all of the appropriate materials. To identify response opportunities, you must use copies of the child's version of the teaching material and, in addition, a teacher's edition. (A teacher's edition alone is not sufficient even though the student materials are reproduced in it. These materials are reproduced with correct answers already marked, making it difficult to know what material the child actually uses.)

Once you have located the child's version of each assignment and its corresponding directions in the teacher's edition, you may begin to identify individual response opportunities. A response opportunity is an occasion for student behavior that may be directly evaluated--reading aloud, underlining, refraining from underlining, coloring, answering questions, etc. Occasions for covert behavior, such as silent reading, that cannot be directly evaluated are not included in a count of response opportunities.

Counting response opportunities in materials designed for a minimum of teacher intervention. As a general rule, each answer representing a child's decision that could potentially be marked right or wrong is counted as a single response opportunity. For example, a typical workbook auditory perception exercise presents several pictures of objects. The child circles the pictures with names beginning with a particular phoneme. Each picture is counted as one response opportunity, since at each picture the child must decide whether or not to circle, and this decision can be directly evaluated. As another example, a typical workbook comprehension exercise presents a picture accompanied by two words or phrases. The child circles the word or phrase that describes the picture. Again, at each picture the child makes a decision that can be directly evaluated as right or wrong, so each picture is counted as one response opportunity. In another example, some workbook exercises may have a different number of response opportunities depending upon the specific directions for use. The workbook page may look like this:



If the child is directed to circle each letter "e," the number of response opportunities is equal to the total number of letters in the exercise, since the child makes a decision at each letter that can be evaluated. If, however, the child is directed to draw a line on the path that has only letter "e's" on it, there are fewer decisions to be evaluated and correspondingly fewer response opportunities counted. In this case, the child either draws a line or does not draw a line on each segment of the path between branches. Thus, there are as many response opportunities as they are between-branch segments of path, i. e., seven. In another typical comprehension format used in later grades, children compose or select answers to questions based on their silent reading of a story. In this case, each question overtly answered counts as one response opportunity.

Counting response opportunities in materials requiring teacher presentation to a group of students. Much new material is presented to students in this manner. All estimates of response opportunities will be derived from analyses of directions in the teacher's manuals, not from actual teacher performance. Teacher's editions of reading materials always make some suggestions on methods of presenting new material, and, in fact, some reading curricula come complete with detailed teacher scripts. Typically, in providing guidance for group instruction, the teacher's edition suggests that the teacher present some new information and then ask the children certain questions, or "elicit" certain responses from them, or ask for a volunteer to perform the newly learned response or to answer questions about the just-read story. If the children are to answer in unison, a response opportunity is provided for each child in the group. However, if the teacher is directed to ask for a volunteer, or call on only one child, only one response opportunity is counted. Since it is a total of all response opportunities that will be calculated, arbitrarily assign this one response opportunity to one child's tally.

Finding the classroom total of response opportunities. When you have located each child's reading assignment and counted the response opportunities in that assignment, simply add these per-child totals to find the classroom total of response opportunities. Some value representing response opportunities must be entered for each child who has a reading assignment, even if that value is zero. Make a note of the number of children involved for use in item 23. On line 19 of the data sheet, enter the classroom total of response opportunities.

20. In computing the value entered on this line of the data sheet, you must closely examine each response opportunity that has been entered in the total for line 19. You will make a judgment about the minimal behavior necessary to achieve a correct answer for each response opportunity, and a second judgment comparing that minimal behavior to the inferred or stated behavioral objective of the item. In order to make such judgments reliably, you must have completed the training keyed to this variable, appropriately determined response opportunities. The following instructions briefly outline the concept of response contingency and review some guidelines for judgments. They are not a substitute for the training of the curriculum experts.

Response contingencies. A student learns what he or she performs. Usually only a small part of the student's activity is public and available to the instructor--that is, a question is answered about material the student has read. In well-designed materials, the student's successful public performance depends upon correct execution of the private act. The question can be answered correctly only if the material has been read. In poorly designed materials, overcueing or inappropriate cueing frequently allows a successful public performance based on trivial private behaviors, as when, for example, a question can be "answered" by merely copying underlined words from the text. In the workbook page illustrated earlier, the "e's" are arranged in a pattern. If the directions are to circle each "e," the responses are overcued, since the pattern as well as letter shape cues the response. If the directions are to draw a line through the path that has only "e's" on it, the response is not overcued.

Group instructional situations are inherently flawed in terms of response contingencies. One child in the group may be responding to the material and cues provided by the teacher, while others may simply remain silent or repeat the first child's response. Because of this contingency problem, an apparently well-designed response opportunity in group instruction may not actually function as such. Thus, we generally recommend that group instruction response opportunities be "counted" as zero. By this we do not imply that children do not learn in these group situations, only that the learning that occurs cannot be simply and directly related to the response contingency aspect of the design of the curriculum.

You may wish to make an exception to this general rule and count group instruction response opportunities in a particular curriculum if: (1) that curriculum provides detailed teacher instruction and training packages to ensure consistent classroom implementation, and (2) that curriculum provides the teacher with instructions about how to minimize the response contingency problem in group instruction.

Guidelines for evaluating private behaviors. Once you have identified the behaviors that will result in a correct answer for each response opportunity, you must judge whether these are, indeed, the behaviors that item was intended to teach. In many cases, behavioral objectives for a lesson or worksheet will be clearly and explicitly identified in the teacher's edition. In other cases, you must infer the objective by examining the lesson or worksheet as well as some of the surrounding teaching material and the general objectives usually stated at the beginning of the teacher's edition.

Once you have identified an explicit or inferred behavioral objective for the response opportunity, and the range of behaviors that result in a correct response for that response opportunity, you may easily answer the question from line 20 of the data sheet, "Can a child respond correctly without behaving in the manner described by the stated or inferred objective?" Ask this question of each response opportunity. Sum the number of "yeses" to this question and enter that number on line 20 of the data sheet.

21. Just as it is important that a correct public response should not be achieved by undesired private behavior, it is also important to learning that the desired private behavior can and does occur. If response contingencies are correct, student error rate provides a direct empirical measure of whether the desired private behavior has occurred. However, it is possible to estimate potential errors from an examination of the curriculum materials. The training keyed to appropriately determined response opportunities

provides guidance in judging whether previous material has prepared the child to perform the correct response. You should complete this training before making the judgments called for in line 21 of the data sheet. If, however, it is possible for you to obtain performance records of the children in a particular classroom on the individual response opportunities you are examining, you may substitute these empirical records for your judgments. (Brief guidelines on the interpretation of errors in children's performances should be included in the training of curriculum experts.)

However, it is most probable that you will simply be making judgments about whether errors are likely on a particular item. Remember that you are concerned only with those response opportunities that are properly cued (those to which you answered "no" in line 20). Answer "yes" to the question on line 21 if the item is obviously ambiguous (as are some of the pictures used in beginning reading exercises to teach auditory discrimination), obviously calls for complex behavior that has not been practiced before (as do some exercises in "just reading aloud" words containing untaught phoneme-grapheme correspondences), or asks the child to respond in an unsupported situation when earlier responses were improperly cued. Add the number of "yeses" to this question and enter this number on line 21 of the data sheet.

22. The number of "yeses" entered on line 20 represents the number of improperly cued response opportunities likely to result in correct answers through undesired behavior. The number of "yeses" entered on line 21 represents the number of undetermined response opportunities likely to result in errors. Subtract each from the total number of response opportunities (line 19) to find the classroom total of appropriately determined response opportunities. Enter this number of line 22 of the data sheet.

23. The number entered on line 22 is the classroom total of appropriately determined response opportunities. To find the average number of appropriately determined response opportunities per child, you must divide this sum by the number of children in the classroom who had reading assignments. (Refer to your calculations for the number of response opportunities [line 19] to find the number of children.) Enter the average number of appropriately determined response opportunities per child on the data sheet and coding sheet.

24. In order to find the percentage of appropriately determined response opportunities per class, divide the number of appropriately determined response opportunities (line 22) by the total number of response opportunities (line 19). Enter this percentage on the data sheet and coding sheet.

Behavioral Analysis of Curricula Coding Sheet

1. Leave blank
2. "
3. "
4. "
5. "
6. "
7. "
8. Enter this number
9. Leave blank
10. "
11. "
12. Enter this number
13. Leave blank
14. "
15. "
16. "
17. Enter this number
18. "
19. Leave blank
20. "
21. "
22. "
23. Enter this number
24. "

Additional Information

Predictive Validity Ratio

Some additional information regarding the predictive validity ratio is presented below. This information can serve as the basis for a letter to or meeting with the teacher to explain this measure and the procedures that must be followed if it is to be estimated in his/her classroom. As indicated previously, obtaining an estimate of the predictive validity ratio requires classroom intervention that may be difficult for the contractor to accomplish. Therefore, it should be regarded as an optional measure.

What is predictive validity? Many curricula claim to adapt to individual differences in children through the administration of diagnostic tests or pretests that are given before any formal teaching of the objectives measured by those tests has taken place. As a result of performance on the diagnostic test, a child either skips or takes a certain segment of teacher material. If the curriculum is to be truly adaptive, the decisions about what teaching material a child is to receive or skip must be accurately made.

There are potentially two types of diagnostic test inaccuracies. If through an invalid diagnostic test, a child is directed to take teaching material that he/she already knows, the child is wasting time that could be better spent on other material. The more detrimental error occurs when a child, through an invalid diagnostic test, is directed to skip teaching material that he/she really needs. The child may never be assigned to this material and, if it is critical to learning later material, may ultimately flounder.

The predictive validity ratio provides a measure of the accuracy of such diagnostic test decisions. It is simply a ratio of the number of accurate diagnostic test decisions to total decisions. Surprisingly, applying this measure to a variety of curricula revealed that many widely used diagnostic tests make invalid predictions.

Measuring predictive validity. In order to identify the accuracy of any single diagnostic test decision, a slight change in the normal classroom routine of prescribing, teaching, and testing is required. Ordinarily, a child takes a diagnostic test, then skips or takes the relevant teaching material and mastery test in that order. The change in routine involves simply following the diagnostic test with the corresponding mastery test (or with a readministration of the diagnostic test) before giving any material designed to teach those objectives. After a child has taken the diagnostic test and its corresponding mastery test (or diagnostic retest), the normal classroom routine resumes. The child takes or skips the relevant teaching material according to the original performance on the diagnostic test.

This slight change in routine is necessary for only one diagnostic test per child. During a selected four-week period, the first time each child takes a diagnostic test, the corresponding mastery test (or diagnostic retest) can be immediately administered, and the data recorded on the Predictive Validity Data Sheet on page B-70. At the end of this time period, the data sheet can be mailed in or collected by the interviewer.

Please note that it is not necessary that each child in the classroom take a diagnostic test in the selected time period. Of course, the more children from whom data are derived, the more accurate the predictive validity ratio will be, but at least ten children from any one classroom will be sufficient. Note also that the names of individual children should not be included on the section of the data sheet returned to the interviewer.

Identifying the diagnostic and mastery tests. Once the time period for data collection is selected, the children themselves will identify the diagnostic tests on which the data are to be collected. The first diagnostic test taken by each child in that time period is included in this study.

Since children will probably be at different points in the curriculum in any one time period, a number of different diagnostic tests will be in the sample. Usually, diagnostic tests are clearly labeled as such in published curriculum materials. Each diagnostic test in the sample must be identified by noting its page numbers on the data sheet. If the classroom uses teacher-developed diagnostic tests, enclosing a copy of each used in the sample would be useful. Should there be any question about whether or not a particular test is diagnostic, the following definition will be helpful: a diagnostic test is one taken prior to any attempt to teach the objectives measured on it and one that has differential consequence depending on student performance (i.e., students either take or skip a particular segment of teaching material).

Mastery tests are often also clearly labeled in published individualized materials. When the published curriculum labels and pairs a diagnostic test with a corresponding mastery test, it becomes simply a matter of ensuring that the appropriately labeled mastery test is administered to each child who has taken the related diagnostic test.

In some curricula, however, mastery tests are not so clearly paired with diagnostic tests. It is sometimes possible to identify a section of a larger posttest that correlates directly with the objectives tested by the particular diagnostic test. This section of the larger posttest may then be administered after the diagnostic test.

Predictive Validity Data Sheet

Teacher ______________________________

Classroom ______________________________

Curriculum in use ______________________________

			Diagnostic Test		Mastery Test	
Student name			Location (page #s or copy)	Recommendation (skip or take material)	Location (page #s or copy)	Performance (pass or fail)
	tear along this line-retain this section	tear along this line-give this section to interviewer				

If there is any doubt about whether such a segment really correlates with the diagnostic test, it is better not to use it in this study. When an appropriate mastery test cannot be paired with a diagnostic test, the same diagnostic test should be readministered and the child's retest performance recorded on the data sheet under the "Mastery Test" column.

If the classroom uses teacher-developed mastery tests, enclosing a copy of each used in this sample would be useful. If the mastery tests used were published in the curriculum, their page numbers should be noted on the data sheet. (If the procedure was "diagnostic test and retest," the same page number should be recorded in both places on the data sheet.)

Time between the two tests. Some time should be allowed to elapse between the two test administrations, especially in the "diagnostic test and retest" procedure. We suggest allowing at least one-half day between the two tests. Beyond that constraint, the amount of elapsed time can be a matter of teacher convenience. However, it is vital that a child not be exposed to teaching material directed at the tested objectives in the time period between the two tests. However, in this interval, each child may use any other teaching material not directed at objectives covered by that child's diagnostic test.

Information That Can Be Obtained from the In-Class Interview

Note: Student name is not necessary for the curriculum expert. We prefer an arrangment that preserves student anonymity, yet does not mix individual student data.

Information for Questions 1 and 9

1. Last diagnostic test taken. Published test or teacher-developed? Title and page numbers if from published text. Copy of test if developed by teacher.
2. Score on this test.
3. Did student skip material consequent to performance on this diagnostic test?
4. Was student prescribed material consequent to performance on this test? Title and page numbers prescribed if from published text. Copy of exercises prescribed if developed by teacher.

5. Fill this out only for those students who skipped material based on the teacher-developed diagnostic test--What material would have been prescribed consequent to a "take" score on this test? Identify page numbers in published text and include copies of teacher-developed exercises.

6. Place a checkmark by those students for whom no diagnosis by means of formal testing has been attempted.

Information for Questions 23 and 24

1. Each child's last in-class assignment (not test).
    a. Page numbers of assignment in the published text.
    b. Copies of assignment, if constructed by teacher.
    c. If some children had a teacher-led group lesson, page numbers in the teacher's edition of the published text.
    d. If tape-and-workbook combination, identify tape so that a transcript can be located by the curriculum expert.

APPENDIX C

VIDEOTAPING

Page

Directions for Producing Videotapes C-1

Directions for Obtaining Measures from Videotapes C-11

Directions for Coding Videotape Information (including Videotape Summary Sheet) C-34

Directions for Producing Videotapes

I. Equipment

There are many systems available that may be used to videotape a classroom, but only one set of recommendations is made in this Appendix. Any comparable products may be used.

Four basic components are necessary. The following list gives the name of the device and a recommended model and/or size.

Video Recording Unit. (Sony, 1/2 inch reel, Model 3600)

Video Camera and Tripod. (Sony, AVC 3200)

Microphone System. (Sony, wireless. A microphone, receiver, and earphone are necessary.)

Video Monitor. (Sony, 9 inch)

In addition, four tapes of one hour each are required for each classroom. This would mean that for 400 classrooms, 1,600 rolls of one-hour videotape are needed. Scotch Color Compatible 1/2 inch tape is recommended.

A portable collapsible cart on which to arrange and carry the equipment to the classroom, and on which the audio equipment can be set up for the taping sessions, is also valuable.

II. Sending and retrieving tapes from sites

Tapes are sent out to the sites twice a year, in the fall and spring. The Field Coordinator should have each tape labeled on the videotape casing as well as on the tape reel. Pertinent information, to be filled in by the persons making the tapes in the classrooms, should include the site, teacher's name, subject area, and date of taping. The labels might look like the following illustration:

SAMPLE LABEL

Site:
Teacher:
Subj:
Date:

It is necessary to send two one-hour tapes for each classroom in the fall and two more in the spring. Accompanying the tapes sent to a site should be a letter to the Site Coordinator and to the teachers, explaining briefly the purpose of the videotaping sessions and the basic procedure that will be used, and emphasizing the confidentiality of the information to be gathered (see page C-3). One or more return-address labels to facilitate the return of the tapes after completion should also be included.

The Field Coordinator must keep a list of the dates on which the tapes are sent out and the dates on which they are received in order to make sure that all tapes are retrieved. A letter may be sent to Site Coordinators and teachers to show appreciation for their cooperation during the videotaping (see page C-4).

III. Videotaping personnel

A team of two people is needed for actual taping. At least one person must be well trained in the use of the videotape equipment (the taper), and the other person should have at least a minimal understanding of how the equipment is used (the assistant). The taper is responsible for setting up the equipment in the classroom, choosing the location from which to tape, and doing the camera work. The assistant is responsible for scheduling taping sessions to match the schedule and convenience of the teacher; explaining the procedures to the teacher; assisting the taper in setting up the equipment; monitoring the audio and video portions of the tape through the use of the video monitor and earphones; keeping time using a stopwatch (to tell the taper when to start and stop a scan); and acting as a liaison between the taper and the teachers in order to make the teachers feel as comfortable as possible.

There are two possible ways of structuring the "team." First, the team may be from the home-base of the contractor and sent out to the various sites from there. If this is the case, more than one team will be necessary. Assuming that two classrooms can be taped in one day by one team (this may be an overestimate), and it is advantageous to complete all taping in two weeks, then 20 teams would be needed for 400 classrooms. The second alternative is to use teams from the sites. The Site Coordinator or school personnel could possibly be used as members of the videotaping team.

Sample Letter to Site to Accompany Tapes

Return Address

Date

Dear Site Coordinator and teachers:

As you know, your classrooms have been selected as part of a study of individualized instruction that is being sponsored by the National Institute of Education. In accordance with the plan for data collection, we will be sending our trained personnel to your school to videotape the classrooms selected.

Videotaping procedures are used to evaluate classroom processes and have been found to be a mobile and reliable data collection device. The information that will be obtained will not be used to evaluate any individual teacher, student, school, or district, and all data will be considered to be strictly confidential.

Enclosed in this package are the videotapes that will be used by our taping team. Each classroom will be taped for one hour of mathematics instruction and one hour of reading instruction. Please store the tapes in a safe, dry place to insure that they are not damaged.

The taping team will contact the Site Coordinator shortly to arrange a schedule that will be convenient for all concerned.

We would like to thank you ahead of time for your support and cooperation in this effort, and to again assure you that anonymity and confidentiality will be maintained.

Sincerely,

Sample Letter to Site after Completion of Taping

Return Address

Date

Dear ______________________:

We have now completed videotaping in your school. This effort was part of an evaluation of individualized programs. In order to assure anonymity, the results of the evaluation will be reported by grade level, not by individual classrooms or by schools. Thank you for your patience and cooperation in this important effort. We especially appreciate the courtesy that you extended to us while we were in your school.

Sincerely,

IV. Specific directions for videotaping in classrooms

The aim of this portion of the Appendix is to give specific directions to the team that will be videotaping in the classrooms.

A. School and classroom behavior of the videotaping team

When entering a school, it is important that any personnel look and act in a professional manner. Following a few simple guidelines will aid in acceptance and facilitate cooperation. This does not imply relinquishing individuality while in the schools, but it does mean that courteous and considerate behavior is expected. These guidelines are suggested:

1) Dress appropriately. Jeans and T-shirts are not acceptable in most schools. If possible, check with the Site Coordinator to determine the mores of each particular school, as they may differ greatly.

2) Be considerate of school parking facilities. Try to unload all equipment at least a half-hour before the students arrive. This will allow the use of convenient doorways. After unloading, be careful not to park in someone's assigned space, or to block other cars or entrances. If the school lot is small, park on the street.

3) When arriving, report to the Site Coordinator and/or office in the school building. Each school has different procedures, but most require that external personnel sign in and sign out.

4) Set up and organize all materials early. Taping during the first class period requires setting up in that room before students arrive.

5) If the teacher's room is used for waiting, leave it in the same way it was found. If there is a coffee maker, leave money (usually 10¢ a cup) for any coffee that you drink.

6) Never talk about a child, teacher, class, or any other school personnel while in the school. Remember that as an outsider, personal views will not be appreciated.

7) Never discuss another school while on site. This can only lead to discomfort of those at the present site, as they will be concerned about what might be said about them at the next site.

8) Remember that the school is a self-contained world with its own values, problems, joys, and conflicts. What is said and done in one world with little fear of misrepresentation cannot always be said or done with similar consequences in the school's world.

9) If any difficulty arises, there should be a back-up system to cover it. For example, there should be a duplicate set of equipment ready in case of an equipment failure. Last-minute scheduling difficulties should be handled by the Site Coordinator or principal if the coordinator is unavailable.

10) Be as efficient as possible. Set up and disassemble equipment quickly and quietly. Try to cause as little disruption to the school processes as possible.

11) Maintain a professional approach at all times. Be friendly but not "chummy." Respect the teacher's position as the authority in his/her classroom. Respect the principal's authority as head of the school. And most of all, respect the enormous load that all school personnel must handle. Try to be flexible and calm no matter what happens.

12) Maintain distance with the children. Being nice does not mean being patronizing or motherly. Do not enter into contact relations with the children (e.g., eye contact with "meaningful smiles," patting), unless it is necessary for the performance of a task.

13) Be open and accepting of any suggestions made by school personnel. If the suggestions do not conflict with the interest in obtaining precise data, try to be as accommodating as possible.

14) If teachers or other school personnel ask questions about the videotaping, be polite, answer briefly, and try not to discuss the matter in any great detail.

15) Be appreciative. Thank all involved for their time and assistance.

16) Remember to sign out or to let someone know you are leaving when you depart from the school building.

17) Always drive slowly and carefully near any school building.

B. Videotaping procedures

1) Setting up

As previously stated, the videotape equipment should be set up in the classroom as early as possible before the actual taping session. If possible, the camera and crew should be in the room for one hour prior to taping so that the teacher and students can become accustomed to the equipment. If two classrooms are to be taped in one day, e.g., one classroom in the morning and one in the afternoon, the equipment can be set up in the first room before the students arrive in the morning and in the second room while the students are at lunch. However, scheduling will not always be so simple, and the taping team must be flexible and able to move the equipment quickly and quietly.

All equipment should be arranged in the clasuroom in a position where it does not inconvenience the teacher or children or block an entrance, but still permits visual and auditory accuracy. Care should be taken in the placement of extension cord wires.

Avoid placing the camera so that it is aimed directly at windows or lights, as this can distort the video portion of the tape and even damage the vidicon tube in the camera. If no other position is available, request that the blinds be pulled so that the camera is not facing into direct sunlight.

The camera should be mounted on a tripod to facilitate smooth movement. Once the equipment is set up, plugged in, and turned on, check and adjust the camera focus and F-stop, and be sure that the camera position permits an unobstructed view of the entire classroom.

Audio reception can readily be checked by having one team member (probably the taper, as the assistant will be monitoring audio signals) travel the full perimeter of the room with the microphone while the other team member checks the range and adjusts the receiver for maximum clarity (using the earphones).

2) Scheduled time

Each teacher should be taped teaching one hour of reading and one hour of mathematics. If the classroom is under open or mixed scheduling (so that reading and mathematics are taught at the same time), two continuous hours should be taped. If a subject is taught in periods lasting less than a continuous hour but there is more than one period of that subject per day, tape both periods.

3) Microphone

Immediately preceding the videotaping session, ask the teacher to put on the microphone. A wireless microphone is preferred, as it is more convenient and unobtrusive, but wired microphones could also be used. In either case, the microphone should be worn in lavalier fashion, usually about eight inches below the teacher's mouth. Request that the teacher remove any jewelry that might clank against the microphone as the teacher moves; pins or necklaces create noises that distort the audio portion of the tape.

The teacher's statements are the primary concern of the audio portion of the tape, and the students' verbal interactions with the teacher are only slightly less important. It is mandatory that a clear, codable record of those verbal behaviors be made, with a minimum of other auditory interruptions.

4) Camera

When the teacher, students, and videotaping team are ready, the taping session should begin with the assistant starting the stopwatch. The taper should focus on the teacher and any student or students with whom he/she is interacting for the first five minutes of the tape. The assistant will tell the camera person when that time is up. The camera then pans across the room for two minutes, focusing on each child for approximately five seconds. The purpose of this scan is to determine whether each child is actively engaged in cognitively related material. It is not necessary (or even desirable) for the person with the stopwatch to indicate each five-second segment; that can be estimated by the camera person. However, at the end of the two minutes, the assistant will indicate that termination. It is important to emphasize at this point that all communication between the two videotape team members must be nondisruptive and preferably nonverbal.

Following those seven specified minutes, the camera should resume following the teacher's movements and should include in the frame the child or children with whom the teacher is currently interacting, if possible. After every five minutes, another two minute scan of the entire class should be made, focusing again on each child for about five seconds. Five minutes of following the teacher is interspersed with two minutes of all the individual students throughout the entire taping session.

5) General instructions

a) The tapeheads on the video recorder should be cleaned thoroughly prior to each day of taping, and between sessions if necessary.

b) When moving the camera to follow the teacher or scan the classroom (called "panning"), be sure to move the camera slowly and evenly; much data are lost through poor camera work, so take time but stay with the subject.

c) If the camera has a zoom lens, change from close-up to full-scene shots infrequently and smoothly. Try to keep the image roughly the same size for most of the taping.

d) It is useful to the person who will observe the tapes if the person operating the camera scans the entire room for a brief period (30 seconds) at the beginning of the taping session, so that a general impression of the room arrangement and size can be obtained.

e) BE CERTAIN THAT THE NAME OF THE TEACHER IS ON THE VIDEOTAPE REEL, AND, IF POSSIBLE, ON THE AUDIO OR VISUAL PART OF THE TAPE AS WELL. This can be done either by making a placard containing the information and focusing on it at the beginning of the tape long enough so that it can be read, or by saying the information into the microphone before giving it to the teacher (be sure the recorder is on!).

6) Checking the tapes

After each taping session is complete, play back several portions of the tape to be sure that both the audio and video portions are recorded clearly. The focus and contrast should be clear, the teacher's and students' voices should be audible with a minimum of static or distortion, there should be no "snow" or other visual distortion, and the identifying information for the teacher, site, subject, and date should be on the tape casing, reel, and, if possible, on the tape itself. If the tape is unusable, schedule retaping as soon as possible.

V. Mechanics of quality control

As soon as all tapes are completed for one site, they are to be sent back to the Field Coordinator using the box in which they were sent and the return address label provided. As soon as they are received, the Field Coordinator should check each tape from beginning to end. The purpose of this preview is to assure codable technical quality. In other words, the tapes must be clear both in sound and image.

Once previewed, acceptable tapes should be checked off on the master list and the date they were received should be noted. If some tapes are unacceptable, the Field Coordinator should note the reason(s) why they are unacceptable. The poor quality tapes should then be erased and sent back to the site with a cover letter explaining the difficulty and requesting a new taping session as soon as possible. The date that any such tapes are sent back to the sites should also be noted on the master list.

Directions for Obtaining Measures from Videotapes

I. Collecting information from videotapes

This portion of the Appendix is of primary concern to the videotape observers and those who will train them, and includes five parts: time unit, observation unit, guidelines for categorization, directions for observing students on task, and guidelines for ratings. To observe a videotape, the video recorder and video monitor should be set up in a stable position at eye level in a quiet room. A clipboard with coding sheets can be held on one's lap or put on the side-arm of a chair. In this position, if one is right-handed, the stopwatch can be held in the left hand and a pencil in the right. The stopwatch is simply a time-keeper and should not be stopped unless the tape is also stopped.

A. Time unit

The Videotape Observation Sheet (see page C-12) is divided into five blocks, each block representing a one minute time period. The one minute block is then further divided (dotted lines) into fifteen second time segments for each observation unit. The teacher's verbal behavior as well as the student's response will be watched for a fifteen second time segment. In the five seconds immediately following the fifteen second segment, the actual recording takes place. After watching the tape for fifteen seconds, the observer marks the appropriate categories either during the fifteen seconds (this becomes easier with experience) or during the five second break.

B. Observation unit

There are three observation units that are used to denote with whom the teacher is interacting. The first symbol, "I," is for an individual student. "Gs" refers to a small group, which can be any size from two to one child less than the whole class. (This does not include children who leave the classroom for any reason, who are set aside from the body of the class for discipline, or absentees.) "G all" means that the teacher is talking to the whole class. The observer must decide whether the majority of each fifteen second block of time is directed towards an individual, a small group, or the whole class.

C. Guidelines for categorization

Each observation sheet has twelve categories across the top of the sheet. The observer must be completely familiar with the definitions and examples of each category prior to observing, since he/she must make continuous, reliable decisions. The following is a list of symbols used and their meanings.

Coder ____________________
Teacher ____________________
School ____________________
Grade ____________________
Subject ____________________
Date ____________________

Videotape Observation Sheet

To Whom Directed	–	M	CM	O	CS	CQ	RES	SIL	TIM	ERR	CHA	Notes
I												
Gs												
G all												
I												
Gs												
G all												
I												
Gs												
G all												
I												
Gs												
G all												
I												
Gs												
G all												

1) - = negative

A negative statement may be either management oriented, personal, or cognitive. It may be stated in the form of a question, exclamation, imperative, or declarative. Any statement is considered negative if the quality of what is said is essentially punitive. Sometimes it is not the words that convey the punishing nature, but the tonal quality used by the teacher. "You did very well on that work," if said in a sarcastic and degrading manner, may mean that the teacher is telling the child how terrible he/she is doing.

In a sense, negatives are the opposite of praise and are seen as negative motivators. Many negative statements involve the demand to stop a behavior; for example, "Sit down," "Stop it," and "You weren't listening." Referring to repetition is another common use of a negative: "How many times do I have to tell you what to do?" Others may imply a negative aspect of the child's overall performance: "You never pay attention," and "Don't you ever read the directions first?"

The negative column is checked whenever a negative occurs, although only one check is permitted for a fifteen second block. The observation sheet on page C-14 shows some samples of possible negative patterns. The first minute shows that the teacher made a negative management statement to an individual child. He/she did the same thing during the second minute even though he/she was working with a small group in cognitive material for the majority of the time. During the third minute, the teacher was working with an individual child or children but made a negative management to the whole class during the first fifteen seconds. The fourth minute indicates that the negative was cognitively related.

2) M = management

Management is any statement or question made by the teacher to an individual child or children that contains no cognitive information but that serves a managerial function only. It may concern discipline or personal functions. Some general examples are: "Yes, you may go to the bathroom," "Please sit down," and "Get your pencil sharpened."

Coder ____________________
Teacher ____________________
School ____________________
Grade ____________________
Subject ____________________
Date ____________________

Videotape Observation Sheet

(Examples of Negative Behavior)

To Whom Directed	–	M	CM	O	CS	CQ	RES	SIL	TIM	ERR	CHA	Notes
I	✓	✓			✓			✓				
Gs												
G all												
I	✓	✓										
Gs						✓✓✓	✓✓✓					
G all												
I		✓		✓		✓✓	✓✓					
Gs												
G all	✓	✓										
I	✓				✓	✓✓	✓✓					
Gs												
G all												
I												
Gs												
G all												

3) CM = cognitive management

Cognitive management is any statement or question made by the teacher that is managerial but cognitively related. It may deal with curricular materials or preparation for a task; for example, "Turn to page 10," "Has your work been checked yet," and "Get your counting sticks out for that problem." If "good" or "right," etc., occurs by itself as the teacher looks at a child's work, it is considered a cognitive management statement. If "good" or "right" occurs in conjunction with specific feedback, it is then considered a cognitive statement (category 5); for example, "Good, you did these division problems correctly."

4) O = other

In the preceding categories, a check was marked if any such behavior occurred during a fifteen second block of time. In this category, a check is marked only if the majority of the fifteen seconds was spent in neither cognitive nor management interactions. This includes personal comments and tangential information not specific to the materials. Examples of this type of verbal behavior are: "How is your brother," "You have a dog, too, don't you," and "Do you ever help around the house like John is doing in this story?"

5) CS = cognitive statement

This type of verbal behavior should be marked if it occurs at any time within the fifteen second time block. A CS occurs when the teacher makes a cognitive statement that does not elicit a response from a child or children. It may be a statement that requires merely an echo response from the child, asking a rhetorical question, lecturing, giving a command, or giving cognitive feedback; for example, "2 + 2 = 4, what is 2 + 2," "Read the sentences, then fill in the blanks using your new vocabulary words," and "If 4 jumps of 3 spaces each = 12, then 4 times 3 = 12." It is important to emphasize here that even a question regarding cognitive material is considered a cognitive statement if the child is not expected to respond, or in fact does not respond.

6) CQ = cognitive question

Mark this symbol on the observation sheet if, during the fifteen second block, the teacher was interacting with a student or students when a response was elicited. This includes asking a direct cognitive question such as "How much is 2 + 2," or making a statement such as "Read me the sentence that tells what Jane is doing." It may also include asking the child to read the directions of what he/she is supposed to do.

Another type of cognitive question may at first appear to be managerial (M), but cognitive information is clearly being passed on; for instance, "How many pages did you do today," "How many were you supposed to do," "How many are you going to do today," and "Did you do more or less than you had planned to do?" Questions that give detailed discussion about the titles of books, how to spell them, and how to write them out, although they are teaching children skills not normally taught in the curriculum, are considered cognitive questions or statements.

7) RES = response

This category is checked on the observation sheet if a child or group of children either initiated a cognitive question or statement, or they responded to one. If it was child initiated, circle the check; if not, just check the appropriate block. Responses may occur in the "I," "Gs," or "G all" rows, depending on who is responding. "Gs" and "G all" responses are usually choral. In order for a response to be child initiated, it must be clear and obvious such as a child going up to the teacher to ask a question. Raising one's hand is not necessarily enough of a signal.

8) SIL = silence

If, during the majority of the fifteen seconds, the teacher or children were not speaking, it is counted as a silence. This may occur while the teacher is checking work, the children are doing seatwork, etc.

9) TIM = time

If the teacher interacts with an individual for 60 seconds in cognitively related material, or if the teacher has three consecutive checks in the "CS" column with either a small group (Gs) or the whole class (G all), then the time category is checked in the fifteen second block at the end of the 60 seconds. In other words, if the teacher is working with a group, a "time" is checked if he/she is lecturing, which is counted on the observation sheet after three consecutive "CS" checks. If, on the other hand, the teacher is working with an individual student, the checks may occur in the "CS" or the "CQ" column, and the interaction is a tutorial one.

10) ERR = error

This box is marked on the observation sheet when the teacher makes an error that goes uncorrected. The error may be factual, as in "2 + 2 = 5," or it may be an extremely noticeable pedagogical error (that is, an error that reflects a very poor teaching technique). An example of this type of error is "What word sounds like 'Mike'," where the expected response is "five." In this case, both words have the long "i" sound but the words themselves do not sound alike, and the child may be more confused by the error than he/she was before the question.

11) CHA = change

A check mark (✓) is made in the change column of the observation sheet when the single student or group of students at the beginning of a fifteen second block is different from the one(s) with whom the teacher was interacting at the beginning of the previous time block. Changes are only counted when the interaction relates to cognitive material. The observation sheet on page C-18 shows an example of a teacher who was working with Johnny during the first block of time and for half of the second block. He/she moved to Judy during the second block and worked with her during the third. The change was noted during the third block of time since at the beginning of the second fifteen seconds he/she was still with Johnny.

If the teacher is working with a small group, the number of students in that group is noted in the space to the right of the change column under "Notes." On the sample observation sheet on page C-18, the teacher is working with a small group containing four students during the second minute.

Coder ______________________
Teacher ______________________
School ______________________
Grade ______________________
Subject ______________________
Date ______________________

Videotape Observation Sheet

(Examples of Changes)

To Whom Directed	–	M	CM	O	CS	CQ	RES	SIL	TIM	ERR	CHA	Notes
I					✓	✓ ✓	✓ ✓				✓	Johnny Judy
Gs												
G all												

To Whom Directed	–	M	CM	O	CS	CQ	RES	SIL	TIM	ERR	CHA	Notes
I							✓					
Gs		✓				✓ ✓	✓					4
G all												

To Whom Directed	–	M	CM	O	CS	CQ	RES	SIL	TIM	ERR	CHA	Notes
I												
Gs		✓			✓	✓	✓	✓			✓	3
G all												

To Whom Directed	–	M	CM	O	CS	CQ	RES	SIL	TIM	ERR	CHA	Notes
I						✓ ✓	✓ ✓				✓	
Gs												
G all		✓										

To Whom Directed	–	M	CM	O	CS	CQ	RES	SIL	TIM	ERR	CHA	Notes
I												
Gs												
G all												

During the third minute, the teacher began working with another small group, and the change was checked in the first fifteen second block as well as the number of students in that group, which was noted under "Notes."

12) Notes

The "Notes" column is used to record the number of children in a small group as well as anything else that may be relevant to a more complete understanding of the tape for future analysis of the data. The observer may use this section to quote an ambiguous comment or possible error so that he/she can go back and look at it again or ask for assistance in determining the proper notation to be used.

D. Directions for observing students on task

During the taping session, the camera will focus for two minutes on the classroom rather than on the teacher, scanning each individual child for about five seconds. The scan of child by child will occur every five minutes through the entire tape. The normal observing of the teacher's verbal behavior will continue during this time, even though the camera is not on him/her. However, a special notation will be made of the children during the two minute scan to determine the percentage of children on task. It should be remembered that the purpose of this measure is to estimate the opportunity that children in a classroom have to learn the information presented. The definition is, therefore, very restricted. It does not mean that the activities that are recorded as off task are not valuable and important.

Definitions

On Task: child actively engaged in work with materials that directly relate to the acquisition of subject-matter-related information.

Clearly On Task: child sitting at a desk or table or on the floor apparently reading, writing, or counting objects.

Clearly Off Task: child playing with blocks, sand, toys, or another child. Child moving from one place to another. Child sitting with hand raised or otherwise "waiting."

Ambiguous: if a child is talking with another child, try to see if the two children are clearly working on an academically related task. If so, consider it on task; if not, consider it off task. If a child is "getting ready" to work by opening up a book or arranging a task, consider it off task unless the child begins to work and stops arranging. If the teacher is "lecturing" and the child is obviously listening or trying to follow some set of directions, consider it on task. If the child is looking around or staring off into space, consider it off task. Any child who is trying to respond to a direct or group directed question by the teacher is considered on task (even if this includes hand waving and shouting out answers).

The notation for each two-minute section should be as follows:

First Pass	Second Pass	Third Pass	Fourth Pass
X	X		
O	X		
O	X		
O	X		
X	X		
X	O		
X	X		
X	X		
X	X		
O	X		
O	X		
O	.		
X	.		
X	.		
X			
O			
O			
X			
X			
X			
X			
X			

An "x" means a child was on task. An "o" means off. If two or more children are together at once, make a decision for each child in the group.

E. Guidelines for ratings

At the completion of observing a videotape, the observer must fill out a rating scale (see page C-22). There are twelve rating categories with a scale of 1 to 5. It is essential that the observer be completely familiar with all of the categories and their definitions prior to watching a tape. Although the rating scale form is not filled out until the observer is finished viewing the tape, he/she may take notes in order to remember the pertinent details for rating each teacher. It is also important to know the categories well so that when an event occurs, such as modeling a response, the observer will be aware that it is a piece of information he/she will need to use later. Most of the ratings deal with the frequency of the occurrence of a behavior, and, therefore, the scale is from "1" (never) to "5" (frequently). Two of the ratings, however, do not require a frequency estimate; for management and tutoring, a quality judgment is required. The following is a list of the twelve categories and their definitions:

1) Active responses sought

The teacher elicits active responses from the students. That is, students are required to verbalize, manipulate, write, or perform some other active form of behavior when the teacher interacts with them as opposed to passive behavior such as listening.

2) Teacher models responses

The teacher very clearly shows the child what a response consists of. For example, the teacher says, "3 + 4 = 7. Write the seven in the box." He/she then shows the child how to do it, and stands and watches as the child does the next problem.

3) Teacher refers to earlier curricular information

The teacher mentions, during an interaction with a student, information that was previously learned; for example, "You sounded out words like this before." It is some communication to the child that the task the he/she is facing is something not totally new.

4) Teacher focuses child's attention on task

This may be judged not only by what the teacher says or does, but also on what the child is doing. That is, does the child look off into space or at the materials? Does the teacher let the student wander off on a tangent? Does he/she get the student back on task?

Videotape Ratings

Teacher ____________________
School ____________________
Subject ____________________

	Frequently				Never
1. Active Responses Sought	5	4	3	2	1
2. Teacher Models Responses	5	4	3	2	1
3. Teacher Refers to Earlier Curricular Information	5	4	3	2	1
4. Teacher Focuses Child's Attention on Task	5	4	3	2	1
5. Teacher Solicits Child's Opinion on Correctness of Responses	5	4	3	2	1
6. Teacher Refers to Earlier Success	5	4	3	2	1
7. Teacher Uses Contingent Praise	5	4	3	2	1
8. Teacher Uses General Praise	5	4	3	2	1
9. Tutoring	5	4	3	2	1
10. Management	5	4	3	2	1
11. Feedback	5	4	3	2	1
12. Teacher Interacts with Individual Students	5	4	3	2	1

5) Teacher solicits child's opinion on correctness of responses

The teacher asks the child to determine whether his/her response is correct.

6) Teacher refers to earlier success

The teacher mentions something the child has done well in the past; for example, "You really did well with multiplication yesterday; division is the reverse process. See if you can figure out how to do these problems. You have done this kind of thing very well in the past."

7) Teacher uses contingent praise

The teacher praises a student for work that is well done; for example, "You added these two numbers correctly. Good." Or, it may simply be telling the child that he/she consistently did something correctly. It is contrasted with the next category, No. 8.

8) Teacher uses general praise

The teacher praises a student in a nonspecific way: "You are working well," "You have done good work," and "You are in a great frame of mind today." The child must interpret what is being praised. No. 7 is specific in relation to the behavior and No. 8 is general.

9) Tutoring

This judgment is based on the clarity, conciseness, and accuracy of the teacher's tutorials. That is, is he/she explaining things clearly, understandably, and logically? Is he/she to the point? Does he/she answer the child quickly and flexibly? Does the child seem to follow the explanation and use the materials appropriately? If the student doesn't understand, does the teacher take a new approach, become more concrete, or find some innovative way of expressing the thought?

10) Management

This judgment is based on how well the classroom works, not necessarily on how low the noise level is. This category was developed because it was clear that despite the fact that teachers have very different management procedures, some being quite constrained and strict and others being quite loose, within each type there are very different successes. That is, there are classrooms where the teachers are very structured and want a quiet room but are not always telling the students to be quiet. Equally, there are classrooms that are open, free, and quite noisy and in which the teacher seems to be quite comfortable. What is basically being measured is the degree of harmony between what the teacher seems to want and what the children seem to be doing. Is the teacher fighting whatever is going on in the classroom? Do the children seem to know what is expected of them? Do they get materials easily? Is everyone functioning individually as well as in a unit? Is the process smooth and even without major disruption?

11) Feedback

This judgment is based on the clarity and frequency with which the teacher gives corrective information to a student about that student's academic behavior. The information should be given in such a manner as to encourage the student to use the information to alter his/her behavior. In other words, the teacher gives the student information with which the student is able to make correct responses in the future. This type of teacher behavior usually occurs in tutorial situations with an individual student during which the teacher questions, the student responds, and the teacher gives feedback on that response.

12) Teacher interacts with individual students

This judgment is based on the degree to which the teacher interacts with individuals within the class as opposed to dealing with small groups or the entire class. In other words, in a strictly individualized classroom where the teacher only interacts with individual students, the rating would be "5." In a strictly traditional model where the teacher deals with the class as a whole, the rating would be "1."

II. Training videotape observers

To assure clear perception and reliability among observers, the training of those observers should be done in a highly systematic way. Only one trainer is necessary, but that person must be fully familiar with the observation schema prior to beginning training sessions for others. The trainer should also have built up his/her own intrarater reliability by that time to assure clarity in the training and ability to answer questions about items that could initially appear to be ambiguous. Both the trainer (who may also function later as one of the observers) and the observers can be trained by the same methods, and all will be referred to as observers in the following description.

Each step in the training procedures that follow is explained in general with specific suggestions for activities. Each session is expected to last approximately one hour.

A. First session: Introduction

The first session should be an introduction to the observation of videotapes. This session should include at least the following:

1) An explanation of how and why the videotapes were collected.

2) A brief description of what the observers will be looking for on the tapes, and how the training process will proceed.

3) Copies of the instructions for observing tapes, to be given to each trainee and read by him/her.

4) A brief explanation of each category. Try to avoid "What if . . ." questions in the first session. If there are questions regarding the substance of a category, answer them to the best of your ability, but do not attempt to cover all possibilities in the first session or confusion is likely to result.

B. Second session: Equipment and time

For the second session and all following, a full set of equipment for viewing videotapes will be necessary. This includes: a video recorder, video monitor, electrical connecting cords (interequipment), take-up reel, and a videotape of a classroom. Each observer should have his/her own stopwatch and clip board.

1) Explain and demonstrate how the machinery is set up and connected. Have the observers examine the machine and each trainee set it up.

2) Demonstrate the threading of the tape onto the take-up reel, and operate the control switch to start, stop, and rewind the tape. Each trainee should also attempt this until he/she has mastered the procedure.

3) Give each observer a stopwatch and demonstrate how to start, stop, reset, and wind it.

4) Give each trainee a blank piece of paper and a pencil. All observers should now watch the tape for timing. Everyone should start his/her stopwatch when the trainer says "start." Tell the observers to make a check on the paper after every fifteen second block when you say "mark." They then have five seconds between each block of fifteen seconds in which to mark their paper. Continue this way for three minutes. Then have the observers mark their papers independently for the next three minutes (do not announce when to "mark").

At the end of the three minutes, have everyone stop again and count their marks. Each observer should have nine checks on his/her paper. This exercise is intended to familiarize the observers with the process of using and estimating the fifteen second intervals, and marking the papers (later, observation sheets) within the five second interval.

C. Third session: Management and cognitive statements

The third session is meant primarily to distinguish between teacher comments that are managerial and those that are cognitive. In this session, you will not be concerned with the subdivisions of each group (that is, M vs. CM, CS vs. CQ, etc.), nor should you try to distinguish to whom the teacher directs his/her comments (that is, I, Gs, or G all). The equipment for this session includes the videotaping equipment from the last session, and a tape of a classroom. Each observer should have a paper with two columns drawn on it, one column marked "Management" and one marked "Cognitive."

1) Discuss the difference between a teacher's managerial statement and a cognitive statement. Emphasize that management statements usually deal with a student's functioning (e.g., "Has your work been checked," "Get a pencil," and "Please sit down."). Cognitive statements usually deal with subject matter (e.g., "What is 2 + 2," and "Read the sentences carefully.").

2) Begin to play the tape as in the previous session, telling the observers when to start (for synchronization purposes). Continuing for three minutes, tell the observers how to mark the columns (tell them whether what occurred in the time segments was managerial or cognitive), for each fifteen second time block. If both a managerial and a cognitive statement occur in a fifteen second time block, indicate that both columns are to be checked for the interval.

3) Watch the tape for another three minute time span, but this time the observers should mark the papers at the appropriate times individually, without prompting from the trainer.

4) Discuss the notation of the last three minute time segment. Correct any errors and clarify ambiguities.

5) Have the observers watch and record their observations for another three minute segment by themselves, recording management and cognitive statements on the observation sheet.

D. Fourth session: Differentiating management and cognitive management

Follow the same procedures as in the third session (that is, discuss the definitions of the terms, give examples, make the distinctions clear, have the entire group observe together, have the observer trainees observe independently, check the observation sheets together and discuss errors and ambiguities, and repeat a three minute session of independent observing). The videotape may be replayed at any time to clarify what was said by the teacher or to show an example of a particular verbal behavior.

E. Fifth session: Differentiating cognitive statements from cognitive questions

Repeat the same procedure followed in the third and fourth sessions (C and D), this time emphasizing cognitives.

F. Sixth session: Combine all managements and cognitives

By this session, all the observers should understand the cognitive and management categories thoroughly. If not, proceed only with those observers who have demonstrated their knowledge of those categories by recording them accurately. For those individuals still having trouble differentiating the two kinds of verbal behaviors, separate sessions should be held to deal with their difficulties.

For those who are prepared for this sixth session, prepare papers on which there are four columns of categories to record: M, CM, CS, CQ. Follow the identical training procedure, using a videotape as in the previous sessions, but increase the observing times to five minutes. After the trainees have completed the second independent observing block, check the rate of agreement in each of the four categories. This is done by each observer summing the number of checks in each of the categories and comparing them with the trainer's sums. For example:

Trainer				Observer			
M	CM	CS	CQ	M	CM	CS	CQ
3	5	1	9	2	6	1	10

Percent Agreement:

M	CM	CS	CQ	
$\frac{2}{3}$	$\frac{5}{6}$	$\frac{1}{1}$	$\frac{9}{10}$	
67%	83%	100%	90%	→ average = 85%

Any percentages that are below 80 are areas of concern. The average percent agreement will indicate which observers are having the most difficulty, and may require additional training.

G. Seventh session: Negatives

Proceeding in the exact format as session F, now add a "Negative" column to the observation sheets (you will now have five columns reading -, M, CM, CS, CQ). It is possible that no negative statements will be encountered in the tape, but an awareness of negative verbal behavior should be emphasized.

H. Eighth session: Other and silence

Gradually increasing the range of categories that the observers must keep in mind while viewing videotapes, proceed as in previous sessions, with emphasis on the new categories ("Other" and "Silence"). Keep in mind that these categories are marked only if they take up the majority of a fifteen second time block. Extra practice may be necessary to master this distinction.

I. Ninth session: Responses

Proceed as in previous sessions, but increase coding times to seven minutes.

Due to the format in which taping is done, each observing sequence will now probably run into or through one of the two minute scans of the entire classroom (these occur throughout the tape at five minute intervals). Explain clearly to the trainees that at this time, their primary concern is with the teacher's verbal behavior, not with the change of camera focus.

Clearly explain the two types of response behaviors: Simple responses are checked (✓), but child-initiated responses are checked and the check is circled (✓⃝).

J. Tenth session: Time and error

As with negatives, an error may not occur during this session. When either an error or a negative does occur, however, it should be noted and replayed so that all observers have a thorough understanding of the occurrence. Continue this session in the previously cited format.

K. Eleventh session: Change and notes

Add these last two categories and proceed as before.

L. Twelfth session: To whom directed

Begin using the observation form (see page C-12) and proceed with the training session as before. The observers will now have to decide to whom the teacher's verbal behavior is directed for the majority of the fifteen second time unit.

M Thirteenth session: Practice I

With this session begin to try to develop reliability between the observers and trainer. This is strictly a practice session. Begin by filling out the identifying information requested in the top right hand corner of the Videotape Observation Sheet. All observers should start their stopwatches at the same time, and then observe at least fifteen minutes of tape independently (everyone can be together in one room). Break for a discussion of any problems and then observe for another fifteen minutes. Where the trainer-trainee agreement is severely off, private sessions may be necessary to correct the observer's errors.

N. Fourteenth session: Practice II

This session is basically a repeat of Practice I. However, when observing is finished, introduce the Videotape Summary Sheet (see page C-35) and have the observers fill out the first 24 items for the 30 minutes' total time they have just observed.

O. Fifteenth session: Practice III

Repeat the Practice II session. Check the percent agreement in each category using the summary sheet information of each observer and comparing it with the trainer's summary sheet. Each observer should calculate his/her own average percent agreement.

At this point, some observers may be finished with their training while others may need more time to practice observing or to backtrack to a category with which they are having difficulty. Proceed to the next step as the observers reach an acceptable reliability level.

P. Sixteenth session: On task

The purpose of this session is to familiarize observers with the observing of the two minute classroom scans throughout the videotapes. This recording on observation sheets should be done simultaneously with the recording of the teacher's verbal behavior. Find a place where such a two minute scan begins just prior to beginning the training session.

1) Explain and discuss the rules for recording "on task" and "off task" behavior.

2) Start the tape and tell the observers when to make an "x" and when to make an "o" (on task and off task).

3) Advance to the next two-minute scan on the tape. Have the trainees record by themselves, marking an "x" or an "o" for each child.

4) Discuss any errors on the trainees' observation sheets and clear up any ambiguities. Replay and re-record any two minute scan if necessary.

Q. Seventeenth session: On task and verbal

1) Review "on task" recording and have the trainees record their observations on another (new) two minute scan.

2) Go to the next two minute segment and have the observers attempt the simultaneous recording of the "on task" information and the teacher's verbal behavior.

3) Check the observation sheets for errors or ambiguities and make explanations where necessary.

4) Advance to the next two minute scan and repeat the simultaneous recording.

R. Eighteenth session: Practice IV

All the observer-trainees should observe an entire tape now (approximately one hour) for both teacher's verbal behavior and students' "on task" behavior. Check all observers' reliability using the procedure on page C-28.

At the end of these practice sessions, some observers may still need more practice than others to achieve acceptable reliability, and they should be given opportunities to practice on their own as well as with the trainer.

S. Nineteenth session: Ratings

This session should familiarize the observers with the twelve ratings to be completed after watching a videotape (see page C-22 for the rating sheet). Discuss the definitions of each category, explain the rating scale, then watch a videotape for about thirty minutes to extract information for the rating sheet. Rate the classroom together and clarify any problems.

T. Twentieth session: Final practice session

Observe an entire videotape and record all information on the observation and rating sheets. Check interobservation agreement.

Reliability on the ratings is determined by summing the absolute difference between the observer's rating and the trainer's rating for each of the twelve categories, and then dividing the sums by twelve to get the average absolute difference. An acceptable average is anything less than 0.50.

Example:

	Trainer	Coder	Absolute Difference	Average
Tutoring	4	3	1	.67
Management	3	4	1	
Feedback	2	2	0	
etc.				

Some observers may now be ready to begin observing tapes on their own for actual usable data. Other observers may need to have one or more practice sessions to clear up any problems they might have. For those who are unable to record teacher verbal behavior and "on task" behavior at the same time, an alternative is to have them observe the tape twice, each time for different information. The ratings can be done after both observation sessions are finished. This is more time consuming than simultaneous observing, but improves accuracy.

III. Reliability

A. Reliability of observers

Several aspects of the data collection are designed to insure interobserver reliability. The basic training of observers is conducted using a program specifically designed to teach reliable videotape observing. The training program includes step-by-step procedures with frequent checks that insure a thorough understanding of all categories. Built-in checks on the reliability of observing between the trainer and the observer guarantee uniformity of interpretations. Further, continuous checks are recommended during observing of actual tapes for usable (not training-session) data. This is best implemented by having the trainer or a different observer watch again every tenth tape completed and record the information from it. Interobserver reliability should be carefully calculated, and any significant disagreements may then be reviewed and observed together by both persons at a second viewing of the tape.

Checking intraobserver reliability involves having each observer watch again a tape he/she has already watched and recorded. This should be done routinely after every twentieth hour of observing. If tapes are recorded differently by the same person, a second person should watch and record the tape to resolve differences.

B. Teacher stability or generalizability

The stability of classroom behaviors over time can vary for different categories. Psychometric theory assumes no changes in the behavior being measured. However, teachers' behavior can change between the fall tape and the spring tape. Some behaviors are more stable than others. (For example, the difference between teacher cognitive statements and management statements is more stable than a teacher's questioning style.) Other behaviors change systematically at different times of the year. (For example, the need to give detailed management instructions decreases over the year.)

Variables that tend to fluctuate a lot have been eliminated from the observation schema, but those that change systematically over time have been retained. By averaging the systematically changing variables, the discrepancies of measurement can be minimized. Only those variables for which the fluctuations of stability can be effectively minimized or readily explained are included in the observation schema for this study.

Directions for Coding Videotape Information

I. Observation summary

The Videotape Summary Sheet must be filled out for each tape. This form (see page C-35) provides space for all identifying information as well as how long the tape was and a summary for each category. One form is used for mathematics and another for reading.

After the first fall tape has been observed, fill in all the identifying information. How many minutes of tape were observed can be most easily determined by counting the number of minutes filled in on the observation sheets. If an entire hour of tape was observed, there should be twelve observation sheets (each sheet provides space for five minutes of tape).

If less than one hour of tape was observed, the following sums must be adjusted so that they are equivalent to one hour of observation. For example, if 50 minutes of tape is observed and the sum for a given category is 25, the proportion used would be:

$$\frac{50}{60} = \frac{25}{x}$$

This converts to $x = \frac{25 \cdot 60}{50}$. The number derived from this formula would be the one recorded on the summary sheet.

For Negatives, Management, Cognitive Management, and Time, sum those categories that were checked in the individual row. Then sum the small group and whole class rows together and put the totals in the correct categories on the summary sheet in the "fall" column.

Other, Silence, and Error are a straight total across group size unit. In other words, it doesn't matter if any of these three categories occurred when the teacher was interacting with an individual, a small group, or the whole class.

Three of the categories, Cognitive Statement Alone, Cognitive Question, and Response, must be summed within each size group. Therefore, there will be one sum for Cognitive Statement Alone/Individual, another for Cognitive Statement Alone/Small Group, and so on. Cognitive Statements Alone, however, will only be counted when they occur in a fifteen second block that does not contain a Cognitive Question. In the example below the Cognitive Statement Alone is counted for B, but not for A.

A

CS	CQ	RES
✓	✓	✓

B

CS	CQ	RES
✓		

Videotape Summary Sheet

Fall Tape--# Minutes Coded _____
Spring Tape--# Minutes Coded _____

Teacher ____________________
School ____________________
Grade ____________________
Subject ____________________

		Fall	Spring	Total
	Negatives			
1.	Individual	____	____	____
2.	Group	____	____	____
	Management			
3.	Individual	____	____	____
4.	Group	____	____	____
	Cognitive Management			
5.	Individual	____	____	____
6.	Group	____	____	____
	Other			
7.	Total	____	____	____
	Cognitive Statement Alone			
8.	Individual	____	____	____
9.	Small Group	____	____	____
10.	Whole Group	____	____	____
	Cognitive Question			
11.	Individual	____	____	____
12.	Small Group	____	____	____
13.	Whole Group	____	____	____
	Response			
14.	Individual	____	____	____
15.	Small Group	____	____	____
16.	Whole Group	____	____	____
	Child-Initiated Responses			
17.	Total	____	____	____

		Fall	Spring	Total
	Silence			
18.	Total	___	___	___
	Time			
19.	Individual	___	___	___
20.	Group	___	___	___
	Error			
21.	Total	___	___	___
	Cognitive Contacts			
22.	Individual	___	___	___
23.	Small Group	___	___	___
24.	Ratio of Cognitive to Management Statements	___	___	___
25.	Percent of Students on Task	___	___	___
26.	Active Responses Sought	___	___	___
27.	Teacher Models Responses	___	___	___
28.	Teacher Refers to Earlier Curricular Information	___	___	___
29.	Teacher Focuses Child's Attention on Task	___	___	___
30.	Teacher Solicits Child's Opinion on Correctness	___	___	___
31.	Teacher Refers to Earlier Success	___	___	___
32.	Teacher Uses Contingent Praise	___	___	___
33.	Teacher Uses General Praise	___	___	___
34.	Tutoring	___	___	___
35.	Management	___	___	___
36.	Feedback	___	___	___
37.	Teacher Interacts with Individual Students	___	___	___

Child-Initiated Responses are those cognitive responses that are circled on the observation sheet. They will always be individual, so a simple count of all circled responses is all that is required on this item.

Cognitive Contacts/Individual is the sum of the individual change checks plus 1 (in order to take into account the first child contacted). Cognitive Contacts/Small Group is the sum of the number of children in small groups, which is recorded in the "Notes" column on the observation sheet. Therefore, for the sample shown in the second section, page C-18, Cognitive Contacts/Individual = 3, and Cognitive Contacts/Small Group = 7, for the four minutes of tape coded in the example.

The Ratio of Cognitive to Management Statements is derived by adding the number of all Cognitive Statements Alone (including individual, small group, and whole group) to the number of all Cognitive Questions, and dividing by the sum of all Management and Cognitive Management statements.

$$CS + CQ \div M + CM$$

After the two passes, fall and spring, simply total the two columns in the "Total" column for items 1-23. For the "Total" column of Ratio of Cognitive to Management Statements (item 24), again calculate the ratio from the total information on CS, CQ, M, and CM using the formula mentioned above.

II. Scan summary

Using the information retrieved from each two minute scan throughout the tape should result in approximately eight passes per hour of videotape. Each tape is summarized to determine the percentage of students on task. To determine this percentage, sum the "x's" of all eight passes and divide by the total number of observations.

$$\frac{\Sigma X}{\Sigma (X+0)} \cdot 100 = \text{Percent of Students on Task}$$

The percentage should then be recorded on the summary sheet, item 25. To obtain the percentage for the total column, average the fall mathematics percentage with the spring mathematics percentage. Do the same for the reading figures.

III. Rating summary

The ratings within each subject area must also be recorded on the summary sheet, items 26-37, and averaged over the fall and spring passes. Simply sum the two ratings in each category and divide by two, rounding to one decimal place. For example:

Fall Mathematics
Active Responses Sought 5 [✓ 4] 3 2 1

Spring Mathematics
Active Responses Sought 5 4 [✓ 3] 2 1

In this case, the recorded total rating for "Active Responses Sought" in mathematics would be 3.5.

At the end of the year, there should be two summary sheets for each classroom: one for mathematics and one for reading.

APPENDIX D

TEST REVIEWS

Page

Comprehensive Tests of Basic Skills D-1

California Achievement Test D-6

Metropolitan Achievement Test D-9

SRA Assessment Survey D-18

Stanford Achievement Test D-24

Animal Crackers D-27

GENERAL INFORMATION

Title: Comprehensive Tests of Basic Skills

Author: None reported

Publisher: CTB/McGraw Hill, Monterey, CA

Publication date: 1974

Forms: The CTBS achievement batteries consist of seven overlapping levels appropriate for use with the following grade levels:

Level A	Grades K.0 - 1.3
Level B	Grades K.6 - 1.9
Level C	Grades 1.6 - 2.9
Level 1	Grades 2.5 - 4.9
Level 2	Grades 4.5 - 6.9
Level 3	Grades 6.5 - 8.9
Level 4	Grades 8.5 - 12.9

There are alternate forms, S and T, for levels 1-4. Forms S and T are parallel forms in content and format and have been statistically equated. Levels A, B, and C are available only in Form S.

The skills areas and tests on each level appropriate for use are as follows:

Level A

Alphabet Skills
1. Letter names
2. Letter forms

Visual & Auditory Discrimination
1. Visual discrimination
2. Sound matching
3. Letter sounds
4. Listening for information
5. Language

Mathematics

Level B

Reading
1. Letter sounds
2. Word recognition I
3. Reading comprehension
4. Word recognition II

Language
1. Language I
2. Language II

Mathematics
1. Mathematics Concepts & Applications
2. Mathematics computation

Level C

Reading
1. Reading Vocabulary
2. Reading comprehension: sentences
3. Reading comprehension: passages

Language
1. Language expression
2. Spelling
3. Language mechanics

Mathematics
1. Mathematics Computation
2. Math Concepts & Applications

Science

Social Studies

Level 1

Reading
1. Reading Vocabulary
2. Reading Comprehension

Language
1. Spelling
2. Language mechanics
3. Language expression

Mathematics
1. Mathematics Computation
2. Mathematics Concepts
3. Mathematics Application

Reference Skills

Science

Social Studies

Level A is a pre-reading test and assumes no school experience.
Level B assumes approximately one year of instruction.
Level C assumes approximately two years of instruction.

Manuals and other technical aids:

(1) Examiner's manuals (Provided for each level, 60-100 pages each)
(2) Test Coordinator's Handbook (one for all levels, 93 pages)
(3) Technical Bulletin (73 pages)

PRACTICAL CONSIDERATIONS

Costs for Levels A and B

Complete Battery (Reading, Language, & Math) MS	$16.45 pkg/35
Complete Battery (Reading, Language, & Math) HS	9.80 pkg/35
Scoring/student	.60

Costs for Level C

Partial battery (Reading, Language, & Math) MS	$16.45 pkg/35
Partial battery (Reading, Language, & Math) HS	9.80 pkg/35
Scoring/student	.60

Costs for Level 1

Partial battery (Reading, Language, Math and Reference Skills) MS	$16.45 pkg/35
Partial battery (Reading, Language, Math and Reference Skills) HS	10.50 pkg/35
Scoring/student	.60

Administration and Timing

The publishers of the CTBS have taken great care in the preparation of the examiner's manuals appropriate for each level. The manuals are well written, provide explicit directions, and contain all the necessary information for teachers to administer and interpret the CTBS.

All tests are timed. The suggested testing schedule for administering the complete and partial batteries of interest to our study are as follows:

Levels A (144 min.), B (157 min.), & C (170 min.) - 4 mornings
Level 1 (173 min) - - - - - - - - - - - - - - - - - 3 mornings

FORMAT AND LAYOUT

The format for all levels of the CTBS is of exceptionally high quality. The illustrations are clear and accurate, the print very readable, and the layout carefully avoids crowding of items and the confusion of inconsistent movement through rows and columns.

ITEMS AND THEIR COVERAGE

Items for Levels 1-4 of the 1974 Edition were derived from Forms Q and R of the CTBS (1968-1969) and additional items reflecting current trends in curriculum prepared by content specialists and CTB/McGraw Hill staff. Two new subtests, Science and Social Studies, were added to the batteries of Levels C through 4. For inclusion in the battery, an item was subject to the following criteria:

1. Difficulty index in the range from .25 to .90 (mean difficulty for all items is .63). Difficulty indices for all items are reported.

2. Consistent ability to discriminate between high and low scoring students.

3. Point biserial correlation (r_{pb}) greater than .20.

4. No bias against black and Spanish-speaking students as judged by reviewers from these respective minority groups.

The development of Levels A-C began by having teachers critique existing primary level achievement tests. Items for the primary edition (Levels A-C) were then written by teachers and curriculum specialists in the respective content areas. Item tryouts were then conducted on a standard sample of students and, in addition, on a special sample of 3,000 students from schools having a minimum of 90 percent black students. Black and Spanish-speaking educators reviewed the items and deleted those items which discriminated against these respective minorities.

CONSISTENCY

The Technical Bulletin presents the data derived from studies of the internal consistency of each of the Levels A-4. Kuder-Richardson Formula 20 correlations are respectable (i.e., >.75) for almost all subtests but are, as would be expected, higher at the higher levels (1-4).

Pearson product-moment correlation coefficients were computed as a measure of the extent to which successive levels of the CTBS measure the same thing. Adjacent levels (e.g., Levels A and B, C and 1) were administered to the same groups of students at two-week to five-month intervals. The results of this study of interlevel articulation of pertinence to our purposes are as follows:

Adjacent Levels	Range of Interlevel Correlations for Total Battery (5-month interval between testings)
A & B	.59 to .78 (.78 for total)
B & C	.45 to .76 (.76 for total)
C & 1	.52 to .80 (.80 for total)

This indicates that adjacent levels of the CTBS do, to a considerable extent, consistently measure the same thing.

VALIDITY

Content validity of the CTBS for our purposes can be determined by matching the curriculum objectives of individualized programs in our study sample with the process and content objectives published for each of the subtests of the CTBS. These process and content objectives for the CTBS are set forth in detail in the Test Coordinator's Handbook.

The intercorrelation coefficients between total battery scores on CTBS/S and total IQ scores derived from Short Form Test of Academic Aptitude (SFTAA) are reported as follows:

	CTBS			
SFTAA	Level 1	Level 2	Level 3	Level 4
Level 1	.64			
Level 2	.79	.82		
Level 3		.85	.79	
Level 4			.82	.77
Level 5				.79

SCORES AND NORMS

The national standardization and norming of the CTBS was based on a probability sample of 130,000 students in grades K-12 drawn from public and Catholic schools in the 50 states. The norm sample was stratified along the following dimensions:

A. School Type
    1. Greater Cities Public (N=37,650)
    2. Other Public (N=84,567)
    3. Catholic (N=8,742)
B. Geographic Region (U.S. Office of Education regions)
C. Average Enrollment Per Grade
    1. Small (N<385) N=46,108
    2. Medium (N<1,923 + N≥385) N=34,482
    3. Large (N≥1,923) N=50,369
D. Community Type (Urban, Town, Rural, Other)

The ethnic composition which resulted from this sampling procedure is as follows:

1. Black - 16.7%
2. Spanish speaking - 7.9%
3. Nonminority - 74.6%

This represents a slight overrepresentation of minorities as compared to the ethnic composition of public schools as reported by the U.S. Office for Civil Rights in 1970.

Separate norms for large cities available upon request.

ADMINISTRATION AND SCORING

Administration of the CTBS is clearly and explicitly detailed in the Examiner's Manual. No other special training is required.

Test batteries may be hand scored or machine scored. Answer sheets available for use with the CTBS include CompuScan, Digitek, IBM 1230, and Scoreze.

COMMENTS

The CTBS is highly recommended by this reviewer as being, in many ways, a model achievement battery. The tests themselves are superior as far as format, layout, validity, and reliability are concerned. The development of the CTBS has been carefully documented, and all claims for the battery are supported by factual data reported in the Technical Manual.

The following is a suggested schedule for using the CTBS in the proposed study.

	September	May
Grade 1	A^s	B^s
Grade 2	B^s	C^s
Grade 3	1^s	1^t

GENERAL INFORMATION

Title: California Achievement Test

Authors: Ernest W. Tiegs, Willis W. Clark

Publisher: CTB/McGraw-Hill, Monterey, CA

Publication date: 1970

Forms: There are two forms--A & B. The two forms were equated using the equipercentile method.

	Level 1 (Gr. 1.5-2)		Level 2 (Gr. 2-4)	
	Items	Time	Items	Time
Reading				
Vocabulary	92	30	40	13
Computation	24	16	45	27
Math				
Comparison	40	14	72	26
Concepts & Problems	47	17	45	22
Language				
Auding	15	6	--	--
Mechanics	38	10	66	22
Usage & Structure	20	11	25	6
Spelling	20	10	25	7
Total Battery	296	114	318	123

Manual and Other Technical Aids:

(1) Examiner's Manual (87 pp.)
(2) Test Coordinator's Handbook (49 pp.)
(3) Bulletin of Technical Data (occasionally published)

PRACTICAL CONSIDERATIONS

Costs for Levels 1 & 2

Complete battery (Reading, Math, & Language) MS	$18.80 pkg/35
Complete battery (Reading, Math, & Language) HS	9.90 pkg/35
Machine scoring/student	.55

Administration and Timing

All tests on the CAT 70 are timed. The suggested testing schedule for administering the complete batteries for Levels 1 and 2 is as follows:

Level 1 (239 minutes)	4 mornings
Level 2 (280 minutes)	4 mornings

FORMAT AND LAYOUT

The format of the CAT is well planned and appealing. The reading subtests are especially well laid out and uncrowded. Illustrations are clear and type very readable. All answers are multiple choice on all subtests which may cause difficulties for the youngest children on certain subtests. On Math Computation, for example, the child must not only compute the answer, but must then find and mark the correct answer from among four alternatives. In reality, this is a fairly sophisticated test-taking skill. For purposes of machine scoring, having the child code his answers in such a manner is extremely convenient. Perhaps it cannot be avoided.

ITEMS AND THEIR COVERAGE

Content objectives for test items are clearly outlined in the Coordinator's Handbook. As with most achievement tests, CAT item coverage was guided by a review of textbooks in reading, mathematics, and language used in the various states and a study of recommended curricular objectives also sampled from various sections of the country. Items were chosen to reflect "curricular relevance," difficulty, and ability to discriminate between high and low scores.

CONSISTENCY

Kuder-Richardson reliability coefficients reported for math and reading subtests on both Level 1 and Level 2 batteries computed at 4 grade levels (1.6 to 4.6) are all at or above .90, except for Reading Comprehension at the 1.6 grade level.

VALIDITY

The CAT 70 has convincing face validity which is supported by the history of its development. For our purposes, the validity of the CAT can be confirmed by matching curriculum objectives of individualized programs in our study samples with the content objectives for each of the CAT subtests.

SCORES AND NORMS

203,684 students
36 states
P.S. stratified by:
1. Geographic region
2. Average enrollment/grade
3. Community type

Minority group representation was not included in the stratification specifications. The authors, however, assert that there is adequate minority group representation in the sample insofar as minority groups participate in public education in the U.S.

ADMINISTRATION AND SCORING

No special training is necessary for teachers to administer the CAT. The Examiner's Manual for each level is well written and provides explicit instructions. Three testing sessions varying in length from 45 minutes to 1 hour are required to administer the entire battery of both Level 1 and Level 2. Tests may be hand or machine scored.

COMMENTS

Suggested schedule for using the CAT in the proposed study:

	September	May
1.	CAT I^A	CAT I^B
2.	CAT II^A	CAT II^B
3.	CAT II^A	CAT II^B

GENERAL INFORMATION

Title: Metropolitan Achievement Test

Authors: Walter W. Durost, Harold H. Bixler, J. Wayne Wrightstone,
George A. Prescott, Irving H. Balow

Publisher: Harcourt Brace Jovanovich

Publication date: 1971

Forms: Primer (Grades K.7 - 1.4)
- Form F (1971)
- Form H (not available for review)

Primary I (Grades 1.5 - 2.4)
- Form F (1970, 16 pages)
- Form G (not available for review)
- Form H (not available for review)

Primary II (Grades 2.5 - 3.4)
- Form F (1970, 20 pages)
- Form G (1971, 20 pages)
- Form H (not available for review)

Manual and Other Technical Aids:

(1) Teacher's Handbook
(2) Teacher's Directions
(3) Manual for Interpreting (128 pages)

Subtests:

Primer

(1) Listening for Sounds - "39 items measure pupils' knowledge of beginning and ending sounds and sound-letter relationships."

(2) Reading - "33 items measure pupils' beginning reading skills."

(3) Numbers - "34 items measure pupils' understanding of basic mathematical principles and relationships."

(Teacher's Directions [Primer], 1971, p. 3)

Primary I

(1) Word Knowledge - "35 items measure extent of pupils' reading vocabulary."

(2) Word Analysis - "40 items measure pupils' knowledge of sound-letter relationships or skill in decoding."

(3) Reading - "42 items measure pupils' comprehension of written material."

(4) Mathematics -
"Part A: Concepts - 35 items measure pupils' understanding of basic mathematical principles and relationships."
"Part B: Computation - 27 items measure pupils' ability to add and subtract one- and two-digit numbers with no regrouping."

(Teacher's Directions [Primary I], 1970, p. 3)

Primary II

(1) Word Knowledge - "40 items measure extent of pupils' reading vocabulary."

(2) Word Analysis - "35 items measure pupils' knowledge of sound-letter relationships or skill in decoding."

(3) Reading - "44 items measure pupils' comprehension of written material."

(4) Mathematics: Computation - "33 items measure pupils' ability to compute."

(5) Mathematics: Concepts - "40 items measure pupils' understanding of basic mathematical principles."

(6) Mathematics: Problem Solving - "35 items measure pupils' ability to apply knowledge in solving numerical problems."

(Teacher's Directions [Primary II], 1970, p. 3)

PRACTICAL CONSIDERATIONS

Costs* for Primer

MRC Machine Scorable Edition (F,G)	$16.50	pkg/35
NCS Machine Scorable Edition (F)	18.45	pkg/35
Primer Battery (F,G)	11.25	pkg/35
Stencil key for test booklet	1.60	each
Teacher's Directions	.85	
Teacher's Handbook	.60	
Practice Page	.10	

*Effective January 1, 1975

Costs* for Primary I

MRC Machine Scorable Edition (F,G,H)	$16.50	pkg/35
NCS Machine Scorable Edition (F)	18.40	pkg/35
Primary I Battery (Hand scored - F,G,H)	11.25	pkg/35
Stencil key for test booklet (F,G,H)	1.85	each
Teacher's Directions	.85	
Teacher's Handbook	.60	

*Effective January 1, 1975

Costs* for Primary II

MRC Machine Scorable Edition (F,G,H)	$18.25 pkg/35
NCS Machine Scorable Edition (F)	19.25 pkg/35
Primary II Battery (Hand scored - F,G,H)	13.00 pkg/35
Stencil key for test booklet (F,G,H)	2.50 each
Teacher's Directions	.85
Teacher's Handbook	.60

*Effective January 1, 1975

Scoring Costs*

Primer	
Machine Scored (MRC Booklet)	$ .50
Hand Scored	1.75
Primary I	
Machine Scored (MRC Booklet)	$ .60
Hand Scored	2.00
Primary II	
Machine Scored (MRC Booklet)	$ .65
Hand Scored	2.25

*These per pupil costs are effective January 1, 1975.

All subtests of the MAT are timed.

Primer

Recommended Sittings	Part	Administration Time in Minutes
1	Practice Page	10
2	Listening (pages 2 & 3)	20
3	(page 4)	5
4	Reading (page 5)	5
5	(pages 6 & 7)	15*
6	Numbers (pages 8 & 9)	5
7	(page 10)	5
8	(page 11)	15*
		Total 80 minutes

*These parts are not teacher-dictated.

It is recommended that the test be administered to groups of not more than 15 pupils. Two settings may be scheduled for one day provided there is a sufficient break between sittings. The whole battery should not be administered in one day.

Primary I

Recommended Sittings	Subtest	Administration Time in Minutes
1	Word Knowledge	15
	Word Analysis	15*
2	Reading	30
3	Total Math	30*
		Total 90 minutes

*Portions teacher dictated, therefore, estimated times.

It is recommended that the Primary I be administered in at least 3 sittings with a sufficient break between sittings. "Word Knowledge" and "Word Analysis" may be combined for a sitting (with a break between them, of course) and "Reading" and "Total Math" as the second and third sittings. The complete test should not be administered in one day.

Primary II

Recommended Sittings	Subtests	Administration Time in Minutes
1	Word Knowledge	18
	Word Analysis	15*
2	Reading	30
3	Math Computation	18
4	Math Concepts	20*
5	Math Problem Solving	25
		Total 126 minutes

*Portions teacher dictated, therefore, estimated times.

The authors of the Primary II test recommend that it be administered in 5 sittings. Two sittings may be administered in one day if a sufficient break in time is provided between them. The complete test should not be administered in one day.

FORMAT AND LAYOUT

Primer

In general, the format and layout of the Primer battery are appealing. However, page 4 in the "Listening for Sounds" subtest seems a bit overcrowded.

The items are consistently numbered from left to right. Finally, the drawings used in the test are recognizable.

Primary I

The physical layout of the Primary I test is well designed. When there are two or more items in a row, they are numbered from left to right. Otherwise, the items are numbered down the page. The directions and samples are clear and concise.

The drawings used in the test are good and the printing is excellent.

Primary II

The Primary II test is also designed well. The only complaint that might be raised is with the number of items per page in the mathematics subtests.

ITEMS AND THEIR COVERAGE

The development of the test items for the 1970 edition of the Metropolitan Achievement tests began with a "curricular analyses." Five sources of information were used to decide what was being taught in schools in the U.S. These were: (1) textbook series, (2) curricular syllabuses from individual school systems, (3) statements of objectives from various state and national committees..., (4) published summaries of curricular content in certain areas..., and (5) statements by curricular experts... (Manual for Interpreting, 1973, p. 18).

Item difficulty for the Primer, Primary I, and Primary II tests ranged from about .20 to .90. The average difficulty for the items on these tests was approximately .60. The items within each test are roughly arranged according to difficulty, i.e., easiest to most difficult.

An item pool of about 12,000 items was used in the original tryout of the battery (five forms at five levels). About half of this number (6,000) was used in the final version of the battery.

Although the authors specify that preference was given in the item selection procedure to those "items answered correctly by progressively higher percentages of pupils at successively higher grades," a table of means for each test for each grade level of the norming population was not included. This perhaps would be the most practical check of how reasonable the difficulty level of each test and subtest is and whether the selection procedure above proved satisfactory.

Primer

The Primer's three sections test types of skills that are reasonable to expect of entering first graders. There are no items that seem to be biased against any cultural group.

The Listening section is intended to measure students' knowledge of beginning sounds and ending sounds and sound-letter relationships. The Reading subtest requires the pupil to identify letters, to select the word that describes a picture and to choose one of three easy sentences that describes a picture. The Numbers subtest measures a child's ability of counting, measurement, numerical recognition and his/her ability to add and subtract one-digit numbers.

Primary I and Primary II

The division of the reading content area into three subtests is very appealing. Reading primarily consists of decoding ability (word analysis subtest), the development of a good reading vocabulary (word knowledge subtest) and the ability to comprehend a written passage (reading subtest). The items of these subtests cover a wide variety of topic areas and seem to be applicable to children of all backgrounds.

Math content is broken into two parts in the math subtest of the Primary I test, concepts and computation. The concept items test a wide range of mathematical abilities such as counting, knowing ordinal positioning, time, place value, simple word problems, and more - less. All of these concepts are reasonable to test at this grade level. The computation items sample from addition and subtraction problems (one and two digit) and verticle and horizontal format. A few problems with three addends are also included.

Math content in the Primary II test consists of the above mentioned subtests as well as a subtest entitled "problem solving." The problem solving subtest contains word problems. Breaking mathematics into three subtests at this grade level is conceptually satisfying. The item coverage in each of these math areas is thorough. The student is required to perform on a variety of pertinent tasks.

CONSISTENCY

Each subtest "was designed to have an internal consistency reliability of about .90 for normal groups at each singel grade level in the grade range intended for the battery." (Manual for Interpreting, 1973, p.21) Two reliability coefficients were reported as well as the standard error of measurement for each subtest. The reliability coefficients are:

(1) Split-half (odd-even) coefficient, corrected by the Spearman-Brown formula.
(2) Saupe's estimate of Kuder-Richardson Formula 20.

Note that the values for the internal consistency estimates and standard error of measurement are very reasonable for all subtests of the batteries.

Primer Battery (Form F)

Test	r_{ke}*	r_{tt}**	SE_{mes} (Raw Scores)
Listening for Sounds	.93	.91	2.3
Reading (Items 1-28)	.93	.89	2.1
Reading (Items 1-33)	.93	.90	2.2
Numbers	.96	.93	1.9

*Saupe's estimate of KR-20.
**Split-half estimates.

Primary I (Form G)

Test	r_{ke}*	r_{tt}**	SE_{mes} (Raw Scores)
Word Knowledge	.88	.94	1.7
Word Analysis	.90	.94	2.0
Reading	.95	.96	2.2
Total Reading (W.K. & Rdg)	.96	.97	2.8
Total Mathematics	.93	.96	2.4

*Saupe's estimate of KR-20.
**Split-half estimates.

Primary II (Form G)

Test	r_{ke}*	r_{tt}**	SE_{mes} (Raw Scores)
Word Knowledge	.93	.95	2.0
Word Analysis	.90	.93	2.0
Reading	.93	.95	2.3
Total Reading (W.K. & Rdg)	.96	.97	3.1
Math. Computation	.86	.91	1.8
Math. Concepts	.85	.89	2.2
Math. Problem Solving	.88	.92	1.9
Total Mathematics (Comp. + Conc. + P.S.)	.95	.96	3.5

*Saupe's estimate of KR-20.
**Split-half estimates.

VALIDITY

The validity of the tests is only briefly mentioned in the Teacher's Handbook for both tests. However, the Manual for Interpreting logically defends the content validity of each subtest of the batteries of the test. Two specific questions are addressed: "(1) content areas and objectives covered by the tests," and "(2) representativeness of the content and objectives..." (Manual for Interpreting, 1973, p. 25). A third question concerning content validity, the appropriateness of the test on the local level, is left for the consumer to consider. In general, the authors' coverage of this type of validity is excellent.

Correlation data between the subtests of the Primer, Primary I, and Primary II batteries and the Otis-Lennon Mental Ability test are also presented. For the Primer, these were .55 for the Listening subtest, .52 for the Reading subtest, and .63 for the Numbers subtest. For the Primary I battery these ranged from .54 to .65. For the Primary II battery the range was .48-.72 with all but two coefficients equal to or greater than .62. The number of pupils in the norming sample for this data was 4,000 to 7,000.

SCORES AND NORMS

Five major variables were used in selecting the norming population.

(1) Socioeconomic index (median family income and median years of schooling of persons over the age of 24 in the community).
(2) Size of the community.
(3) Geographic region.
(4) Public versus non-public school system.
(5) Mental ability test scores.

The Manual for Interpreting describes in detail the selection and description of the norming sample. In addition, a more detailed statistical description is contained in "Special Reports Nos. 7 and 8" published by Harcourt Brace Jovanovich, Inc.

The standardization took place at two times during the school year, in the fall (October, 1969) and in the spring (April, 1970). Norm conversion tables are contained in each test's Teacher's Handbook.

ADMINISTRATION AND SCORING

No special training other than studying the Teacher's Handbook and Teacher's Directions is needed in order to administer the MAT.

The materials needed for the test are soft lead pencils (No. 2), a "Testing--Do Not Disturb" sign for the door, a watch or clock with a second hand, a test booklet for each student and one for the administrator as well as a copy of Teacher's Directions for the administrator.

For the machine scorable booklets, certain information needs to be coded by the administrator, e.g., student's name, date of birth, and sex.

Hand scoring the MAT is straightforward. However, if a large number were to be hand scored, it could prove to be very cumbersome and time-consuming.

Hand scoring of this or any standardized test must be monitored closely in order to insure accuracy of the results. Quality control for the Evaluation Project (LRDC) consists of a person scoring the test, another checking at least every other page of those tests previously scored, and finally, a check for errors in scoring by the data coder.

COMMENTS

This reviewer found no major flaws in the content and construction of the Primer, Primary I, and Primary II Metropolitan Achievement Tests. It is technically well designed and intuitively appealing.

GENERAL INFORMATION

Title: SRA Assessment Survey: Primary Edition of the Achievement Series

Authors: Robert Naslund, Louis P. Thorpe, D. Weltz Lefever

Publisher: Science Research Associates

Publication date: 1972

Forms: Primary I (Grades 1.0 - 2.5)
  Form E (1972, 8 pages)
  Form F (1972, 8 pages)

Primary II (Grades 2.5 - 4.0)
  Form E (1972, 8 pages)
  Form F (1972, 8 pages)

Manual and Other Technical Aids:

(1) Examiner's Manual
(2) Technical Brief (14 pages)
(3) Technical Report (118 pages)
(4) Using Test Results (76 pages)

Subtests:

Primary I

(1) Reading (43 items)
  Word-Picture Association
  Sentence-Picture Association
  Comprehension
  Vocabulary

(2) Mathematics (53 items)
  Concepts
  Computation

Primary II

(1) Reading (52 items)
  Reading Comprehension
  Vocabulary

(2) Mathematics (58 items)
  Concepts
  Computation

PRACTICAL CONSIDERATIONS

Costs* for Primary I

Primary I Reading	$3.50 pkg/25
Primary I Mathematics	3.50 pkg/25
Practice Sheets	1.00 pkg/25

Costs* for Primary II

Primary II Reading	$3.50 pkg/25
Primary II Mathematics	3.50 pkg/25
Practice Sheets	1.00 pkg/25

Quantity prices: 20-199 pkg, each $3.15
200 or more, each $2.98

*Prices effective August, 1974

Each package contains hand-scorable test booklets/answer sheets for 25 students, one Examiner's Manual, and a Growth Scale Chart. A copy of the Conversion Table Booklet, including a Rights Key, is shipped with each order.

Scoring Costs*

The Primary I and Primary II test booklets are consumable, i.e., they also serve as the answer sheets. In order to use the SRA scoring service, these booklets are rented. That is, for one price per student all of the testing materials (e.g., test booklets, Examiner's Manual, User's Manual, Technical Report, etc.) are provided. In addition, SRA will score the tests. Three types of scoring services are available with the base price of $1.11 per student ($27.75 per class of 25). This includes national norms, grade equivalents, and raw scores. More elaborate analyses can be requested at additional costs.

*Effective August 30, 1974.

The subtests of the Primary I and Primary II are not timed.

Primary I

Recommended Sittings	Part	Administration Time in Minutes (Estimated)*
1	Reading	
	Word-Picture Association	5
	Sentence-Picture Association	5
	Comprehension	25
2	Vocabulary	15
	Mathematics	
3	Concepts	35
4	Computation	20
		Total 105 minutes

Primary II

Recommended Sittings	Part	Administration Time in Minutes (Estimated)*
1	Reading	
	Comprehension	25
2	Vocabulary	15
3	Mathematics	
	Concepts	35
4	Computation	20
		Total 95 minutes

*The times listed are the suggested amounts of time each section is expected to take. The authors suggest a testing period until at least 90 percent of the students have finished.

Two testing sessions a day are recommended.

FORMAT AND LAYOUT

The reading and mathematics sections of the Primary I and Primary II tests are separate forms. This allows for ease in administration.

Overall, the tasks are well designed. The print is clear and the drawings are well done.

However, two criticisms can be raised. First, Form E of both the Primary I and Primary II tests is printed with brown ink. The print on the reading portion of the Primary I test which was reviewed was light which made this subtest difficult to read. The other subtests printed in brown are, on the other hand, easily readable. Second, the weight of the paper on which the tests are printed is too light. Because both sides of the page are used, the print on the reverse side of the pages shows through.

ITEMS AND THEIR COVERAGE

The development of the test items for the Primary tests included: an examination of basal texts which "account for those used in an estimated 75 percent of the classrooms in the United States"; examination of supplementary information supplied by the text publishers; and curriculum guides published by boards of education. These were supplemented by review of other testing programs and a telephone survey of curriculum specialists in the state departments of boards of education, large cities, and smaller cities.

The content of the mathematics subtests of the Primary I and Primary II tests covers a wide and representative range of content in an interesting manner. The division of the content into concepts and computation is reasonable. The concepts section of the Primary I test has four parts: (1) sets and numeration, (2) operations, (3) measurement, and (4) place value. The computation section tests the students in the addition and subtraction of one- and two-digit numbers. The concepts section of the Primary II test has five parts: (1) sets and numeration, (2) operations, (3) measurement and geometry, (4) place value, and (5) problem solving. The computation of the Primary II section tests the students in addition, subtraction, multiplication, and division.

The reading portion of the Primary I and Primary II tests is divided into two sections, comprehension and vocabulary. The comprehension section of the Primary I includes: choosing the one picture of three presented that is best described by a given sentence; reading a short story and answering items that ask for restatement of the material, gathering information and reading beyond the story. The Primary II comprehension section also has items that ask the student to restate what s/he has read and to gather information in a passage. In addition, the student is also asked to summarize and sequence information and draw conclusions from short passages.

There are two faults that this writer has found with the reading portion of the Primary I and Primary II tests. First, the passages in the comprehension subtests are long. If they were shorter, a greater variety of content could be included. The second criticism is more critical. Neither of these tests includes a subtest on decoding, e.g., sound-letter relationships, discrimination of parts of words (beginning, median, ending). The ability to decode, which is essential to reading, probably should be tested in the Primary II test but most definitely should be covered in the Primary I test.

CONSISTENCY

Kuder-Richardson reliability estimates (KR-20) were reported for both forms of the Primary I and Primary II tests at three grade levels. (Reliability estimates were reported for three grade levels because norming data is also reported for them.)

	Primary I		
	Grade 1	Grade 2	Grade 3
Form E			
Reading	.88	.94	.93
Comprehension	.79	.89	.88
Vocabulary	.81	.89	.86
Mathematics	.89	.91	.86
Concepts	.82	.83	.77
Computation	.83	.88	.78

	Primary I		
Form F	Grade 1	Grade 2	Grade 3
Reading	.82	.93	.92
Comprehension	.73	.86	.87
Vocabulary	.72	.89	.86
Mathematics	.87	.89	.90
Concepts	.79	.79	.85
Computation	.80	.87	.82

	Primary II		
Form E	Grade 2	Grade 3	Grade 4
Reading	.91	.93	.94
Comprehension	.86	.87	.89
Vocabulary	.84	.87	.90
Mathematics	.88	.91	.92
Concepts	.83	.86	.87
Computation	.76	.85	.86
Form F			
Reading	.90	.92	.94
Comprehension	.81	.85	.89
Vocabulary	.85	.88	.91
Mathematics	.89	.92	.91
Concepts	.84	.87	.88
Computation	.79	.87	.84

These reliability estimates are reasonable. The estimates for standard error of measurement (SE_{mes}) for raw scores are listed below.

	Primary I			Primary II		
	Grade 1	Grade 2	Grade 3	Grade 2	Grade 3	Grade 4
Reading	2.79	2.58	1.99	2.83	2.70	2.21
Mathematics	3.08	2.81	2.14	3.06	2.90	2.29

These are also reasonable for the number of items on the tests.

VALIDITY

The authors only talk about the content validity of the achievement tests. They state that they relied on curriculum specialists in making decisions about the test content. The final validity decision however, is placed on the test consumers. This information about validity was found in the pamphlet "Using Test Results." This lack of validity information is perhaps the major fault of the tests.

SCORES AND NORMS

The standardization study for the achievement tests was conducted in April, 1971. This study involved usable test results of 155,567 students from 816 schools in 224 school districts, selected according to a three-stage random sampling plan. The first phase of this sampling plan divided the country into nine geographic regions (following the guidelines of the U.S. Bureau of Census) and selected 224 school districts from these nine strata. The second phase of the sampling procedure categorized the school districts randomly selected in Phase One according to six catagories:

(1) All schools
(2) Large city schools
(3) Title I schools
(4) High SES schools
(5) Rural/small town schools
(6) Nonpublic schools.

One school in each category which had a Grade 12 and one school which did not have a Grade 12 was then selected. In the third phase, classrooms were randomly selected from each school.

This sampling procedure allowed norm information to be computed for each of the six catagories listed in Phase Two. Users can request comparisons with any of these populations when SRA scoring is used, a very positive aspect of the test. In addition, norming data is provided for three grade levels for each test.

(NOTE: The above information was extracted from the Technical Report.)

ADMINISTRATION AND SCORING

Administration and scoring of the Primary I and Primary II test seems simple and straightforward.

The materials that are needed for administration are: a practice sheet and test booklet for each student, standard pencils (No. 2), a copy of the test booklet and Examiner's Manual for the test administrator, and a clock or watch.

If the booklets are to be machine scored, certain information must be hand coded by the administrator. If hand scoring is to be used, quality control measures should be provided.

COMMENTS

On the plus side of this test is the unique scoring alternatives which use the excellent norming procedure.

On the debit side is: (1) the absence of validity information and (2) the lack of decoding items in the reading subtests. These are considered major faults by this reviewer.

GENERAL INFORMATION

Title: Stanford Achievement Test

Authors: Richard Madden, Eric F. Gardner, Herbert C. Rudman, Bjorn Karlsen, Jack C. Merwin

Publisher: Harcourt Brace Jovanovich, Inc., New York

Publication date: 1973

Forms: Primary I (Grades 1.5 - 2.4)
- Vocabulary (37 items)
- Reading (87 items)
- Word Study Skills (60 items)
- Math Concepts (32 items)
- Math Computation and Application (32 items)
- Listening Comprehension (26 items)

Primary II (Grades 2.5 - 3.4)
- Vocabulary (37 items)
- Reading (93 items)
- Word Study Skills (65 items)
- Math Concepts (35 items)
- Math Application (28 items)
- Spelling (43 items)
- Social Science (27 items)
- Science (27 items)
- Listening Comprehension (50 items)

Primary III (Grades 3.5 - 4.4)
- Same as Primary II (above)

There are three forms for each level: A, B, and C. Form C is reading and math tests only.

Manual and other technical aids:

1. Teacher's Guide for Interpreting (55 pages)
2. Teacher's Directions for Administering (32 pages)
3. Norms Booklet (24 pages)
4. Stanford Index of Instructional Objectives (21 pages)

PRACTICAL CONSIDERATIONS

Costs* for Primary Level I

Primary Level I Battery (Hand scored, A & B)	$11.25 pkg/35
MRC Machine Scorable Booklet (A, B)	15.95 pkg/35
NCS Machine Scorable Booklet (A)	18.20 pkg/35
Scoring service, per complete battery	.65

Costs* for Primary Level II

Primary Level II Battery (Hand scored, A & B)	$13.60 pkg/35
MRC Machine Scorable Booklet (A, B)	18.25 pkg/35
NCS Machine Scorable Booklet (A)	20.85 pkg/35
Scoring service, per complete battery	.65

Costs* for Primary Level III

Primary Level III Battery (Hand scored, A & B)	$14.50 pkg/35
MRC Machine Scorable Booklet (A, B)	20.75 pkg/35
NCS Machine Scorable Booklet (A)	23.75 pkg/35
Scoring service, per complete battery	.45

*Effective January 1, 1975.

All subtests of the SAT are timed.

Primary Level I --4 hours 10 minutes (4 days recommended)
Primary Level II--5 hours 40 minutes (6 days recommended)

FORMAT AND LAYOUT

The format is crowded in the Reading and Word Study Skills subtests; however, the illustrations and print are of acceptable quality.

ITEMS AND THEIR COVERAGE

Items in the 1973 edition of the SAT reflect the authors' analyses of the most widely used textbook series in the various subject areas, a wide variety of courses of study, and the research literature pertaining to children's concepts, experiences, and vocabulary at successive ages or grades. The major goal of the authors was "to make sure that the content of the test would be in harmony with present instructional objectives and measure what is actually being taught in today's schools." A table of the p values for each item at three grade levels is included in Part II of the Manual. No item was retained having a p value less than .20.

CONSISTENCY

Reliability coefficients reported are greater than .85 for all subtests, except for science and social science (.73 and .69, respectively). Two coefficients of reliability were computed for each subtest:

1. Split-half corrected by Spearman-Brown Formula (r_{11})
2. Kuder-Richardson Formula 20

VALIDITY

Content validity of the SAT for our purposes can be determined by matching the curriculum objectives of individualized programs in our study samples with the instructional objectives published for each of the subtests of the SAT. Apart from a discussion of the meaning of validity of an achievement test, no data supporting the SAT's claims to validity are presented in the manual.

SCORES AND NORMS

Norms are based on a restandardization of the SAT completed in 1973. This restandardization was undertaken in part as a response to "the significant changes that had occurred in the elementary school curriculum in the intervening years (i.e., 1964-1970)." The norming sample was drawn from 109 school systems in 43 states and finally included over 275,000 pupils. The norm sample attempted to be representative of the national population in terms of:

1. Geographic region
2. Size of city
3. Socio-economic status
4. Public and non-public schools

Blacks comprised 11.6% of the norm sample; Hispanic-Americans, 4.6%, accurately reflecting the proportions of these ethnic minorities in the 1970 national population. More specific data is available in the Technical Data Report, available upon request.

ADMINISTRATION AND SCORING

No special training is necessary for teachers to administer the SAT. The manual of administration is well written and explicit. A practice test is available to familiarize students with standardized test format. A six half-day testing schedule is proposed in the manual. A substantial "Teacher's Guide for Interpreting" is available and is a useful document to help teachers utilize test results in their instructional planning. Machine scoring is available from the publisher.

GENERAL INFORMATION

Title: Animal Crackers, formerly Gumpgookies

Authors: Dorothy C. Adkins, Bonnie L. Ballif

Publisher: CTB/McGraw-Hill, Monterey, CA

Publication date: 1975

Form: There is one form which is designed for individual or small group administration in kindergarten and first grade. The test is composed of five scales and these are:

(1) School Enjoyment
(2) Self-Confidence
(3) Purposiveness
(4) Instrumental Activity
(5) Self-Evaluation

There are twelve items on the test for each of these five scales, although factor analysis results indicate overlap of items over scales. The test booklet consists of squares (six per page) each of which contains an identical pair of storybook-like animals. As the child looks at these pictures, the examiner reads, for example, "This bear (points to animal on left) shows its work to the teacher." "This bear (points to animal on right) hides its school work." Then says: "Now show me your bear."

CONSISTENCY

The reported Kuder-Richardson formula 20 reliability coefficient for first grade samples was .98. Intercorrelations among the components ranged from .76 to .92 for the first grade samples.

VALIDITY

Earlier studies on Gumpgookies (mostly the same items as Animal Crackers) have shown low positive correlations with age and intelligence, and significant relationships between test scores and teacher ratings of motivation. Virginia Shipman (ETS) is using Gumpgookies currently and finds that kindergarten scores predict significantly to third grade achievement scores. From a face validity standpoint the items are clearly stated, making minimal vocabulary demands and seem to cover well the whole area of "school adjustment."

APPLICATIONS

Both SRI and Abt Associates used Gumpgookies in their evaluation studies of Project Follow Through. In the Abt study, there was generally low positive correlation between Gumpgookies and various achievement and affective measures. Abt further found that Gumpgookies scores are not normally distributed, giving evidence of a ceiling effect. The SRI results also indicate some variation in responding as a function of ethnic group membership. CTB/McGraw-Hill has provided norms from a natural tryout of Animal Crackers. The sample size was 10,899 divided almost equally between first grade and kindergarten. This study indicated that black and Spanish-speaking children score lower than others in the test, but these differences are non-significant.

ADMINISTRATION AND SCORING

No special training is required for teachers to administer Animal Crackers. The Examiner's Manual is well written and provides explicit instructions. The test may be administered individually, taking 20-30 minutes per pupil or in small groups (requiring a longer instruction and practice session) taking 45-60 minutes. There are hand-scorable and machine-scorable test booklets. The latter is on light-weight paper and pages would have to be backed with an opaque sheet so that subsequent pictures don't show through.

An Individual Performance Record and a Group Performance Record are available for hand-scoring and analysis. Norms are available for items, components, and overall scores from the CTB/McGraw-Hill national sample.

PRACTICAL CONSIDERATIONS

Costs

Each package of 30 contains an Examiner's Manual; the hand-scorable edition also contains a Group Performance Record with Scoring Key.

Code

ANCR-M/S	Test Books (Machine-scorable)	$12.80 pkg/30
ANCR-H/S	Test Books (Hand-scorable)	11.00 pkg/30

Accessories

Additional Examiner's Manuals	$ 1.50 ea.
Administration Booklets	3.75 ea.
Individual Performance Records	3.50 pkg/100
Group Performance Records	.10 ea.

Scoring

Basic Scoring Service	$ .60/student

Optional Scoring Service

Frequency Distribution	$ .08/student
Administrator's Summary	.03/student
	$25.00 minimum grade & level